DOCUMENT RESUME

ED 195 336 PS 011 846

AUTHOR	Travers, Jeffrey: And Others
TITLE	Research Results of the National Day Care Study. Final Report of the National Day Care Study. Volume II.
INSTITUTION	Abt Associates, Inc., Cambridge, Mass.
SPONS AGENCY	Administration for Children, Youth, and Families (DHEW), Washington, D.C.
PUB DATE	Oct 80
CONTRACT	105-74-1100
NOTE	285p.: For related documents, see ED 131 928-930, ED 147 016, ED 152 114, ED 160 188, ED 168 706, ED 168 733, and PS 011 846-849.
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	Child Development: Classroom Environment: *Day Care: Day Care Centers: Early Childhood Education: Federal Programs: *Measures (Individuals): Minority Group Children: National Surveys: *Policy Formation: Preschool Children: *Research Methodology: *Sampling: Teacher Qualifications: *Test Results
IDENTIFIERS	*National Day Care Study

ABSTRACT

This final report of the National Day Care Study (NDCS), Volume II, provides researchers, social scientists and lay readers with information for judging the soundness of the evidence underlying NDCS conclusions about relationships between regulatable center characteristics and the outcome of care for the child. Thus, Volume II makes free use of the technical apparatus of developmental psychology and statistics. In order to allow this volume to be read alone, without the necessity of constant cross-reference to Volume I: "Children at the Center," certain sections of that volume are included in Volume Two. In particular, the sections of Chapter One that address the study design and variables have been taken substantially from Volume I, as has the portion of Chapter Two that describes the study sample. Other sections of Chapter One and Two are new, including a discussion of general analytic issues and approaches. Chapters Three through Six describe instruments, analyses and results linking regulatable center characteristics to caregiver behavior, child behavior and child test scores. These chapters constitute detailed support for Chapters Five and Six of "Children at the Center," which summarized the study's results on quality of care. The major findings of the NDCS are summarized in the Preface to Volume II. They are restated in Chapter Seven, amplified by details from Chapters One through Six. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

RESEARCH RESULTS OF THE NATIONAL DAY CARE STUDY

PS 011846

ED195336

OVERVIEW OF NDCS FINAL REPORT VOLUMES

Results of the National Day Care Study and its major supporting study, The National Day Care Supply Study, is presented in a five-volume final report. Contents of these volumes are as follows:

Volume I

Children at the Center: Summary Findings and Policy Implications of the National Day Care Study presents in summary form the major findings and implications for federal day care policy of the National Day Care Study, a four-year study of the effects of regulatable center characteristics on the quality and cost of day care for preschoolers. Volume I serves both as a self-contained volume for the policy makers and as the foundation for the detailed presentation of results in Volumes II, III and IV. (Executive summaries of Supply Study findings and findings of an Infant/Toddler Study are included as appendices to Volume I.)

Volume II

Research Results of the National Day Care Study is a companion volume to *Children at the Center*. Volume II documents the analyses and results of the NDCS for the technical reader who seeks a more thorough understanding of the study from a research perspective. Volume II thus provides the quantitative support for the findings and policy conclusions reported in *Children at the Center*.

Volume III

Day Care Centers in the U.S.: A National Profile 1976–1977, the final report of the National Day Care Supply Study, is based on data gathered from a national random sample of over 3000 day care centers, stratified by state. Summary information is presented on characteristics of children and families served, center programs, staff, finances and regulatory compliance. Discussion of results is augmented by over 150 statistical tables.

Volume IV

Technical Appendices to the National Day Care Study is a compendium of technical papers supporting the most important conclusions of the study. These papers form the basis for the summaries in Volumes I and II. NDCS appendices are bound in three sections as follows.

Volume IV-A, *National Day Care Study Background Materials*, contains three papers, each of which establishes a distinctive context for the NDCS: a literature review focused on effects of group care and regulatable characteristics of the day care environment; case studies of the history and current practice of day care in the three NDCS sites (Atlanta, Detroit, Seattle); and a review of child development issues relevant to the NDCS from the perspective of black social scientists.

Volume IV-B, *National Day Care Study Measurement and Methods*, presents individual reports on a series of technical tasks supporting the principal analyses of the effects of key center characteristics on children. Among the topics covered are: analysis of alternative measures of classroom composition; psychometric analysis of the NDCS test battery; and analyses of several other more peripheral instruments used in the study. Also presented are results of a special survey of parents of subsidized children taken during Phase III, analyses of the impact on children of other center characteristics, such as physical space and program orientation, and econometric analyses.

Volume IV-C, *National Day Care Study Effects Analyses*, also a series of individual technical reports, begins with a presentation of the major effects analyses based on the two behavioral observation instruments, and then moves to a detailed treatment of the development and use of adjusted test score gains. The links among caregiver and child behavior, child test scores and other dependent measures are explored. Also detailed are results of the Atlanta Public School (APS) controlled substudy and APS replication substudy.

Volume V

National Day Care Study Documentation and Data gives a brief overview of NDCS data collection instruments and data files. Part A consists of the instruments themselves, including interview and data collection forms, observation systems and cognitive tests. Part B consists of data dictionaries; these describe every variable in the NDCS analytic data files. Part C provides codebooks for the data files. Parts B and C are available on computer tapes, which are readable independent of specific computer systems. Note that computer tapes are available only from Abt Associates.
Copies of the final report may be ordered from:

- EXECUTIVE SUMMARY (ONLY)
  Day Care Division
  Administration for Children, Youth and Families
  Office of Human Development Services
  Department of Health, Education and Welfare
  400 6th Street, S.W.
  Washington, D.C. 20024

- EXECUTIVE SUMMARY, Volumes I–IV
  ERIC Document Reproduction Service
  Computer Microfilm International
  P.O. Box 190
  Arlington, VA 22210

- EXECUTIVE SUMMARY, Volumes I–V
  Abt Associates Inc.
  55 Wheeler Street
  Cambridge, MA 02138

Earlier NDCS publications available from ERIC (hard copy or microfiche) are:

National Day Care Study First Annual Report, Volume I: An Overview of the Study [order number ED 131 928], *Volume II: Phase II Design* [order number ED 131 929], and *Volume III: Information Management and Data Collection Systems* [order number ED 131 930] (Cambridge, MA: Abt Associates, 1976).

National Day Care Study Second Annual Report [order number ED 147 016] (Cambridge, MA: Abt Associates, 1977).

National Day Care Study Preliminary Findings and their Implications [order number ED 152 114] (Cambridge, MA: Abt Associates, 1978).

Final Report of the National Day Care Study
VOLUME II

October 1980

RESEARCH RESULTS OF THE NATIONAL DAY CARE STUDY

Jeffrey Travers
Barbara Dillon Goodson

with

Judith D. Singer
David B. Connell

Editor: Sally Weiss

Prepared for:

Day Care Division
Administration for Children, Youth and Families
Office of Human Development Services
Department of Health, Education and Welfare

Allen N. Smith, Government Project Director

Prepared by:

Abt Associates Inc.
Cambridge, Massachusetts

Richard Ruopp,
National Day Care Study Director

Jeffrey Travers,
Associate Director

Nancy N. Goodrich,
Technical Coordinator

Contract No. 105-74-1100

TABLE OF CONTENTS

	Page
GLOSSARY	VII
FOREWORD	XVII
PREFACE	XIX
ACKNOWLEDGEMENTS	XXIX
CHAPTER ONE: INTRODUCTION	1
Study Objectives	1
Study Organization	1
Phase III Design	4
Variables and Measures	8
Independent Variables and Measures	10
Background Variables	10
Policy Variables: Definitions	11
Policy Variables: Measures	11
Dependent Variables and Measures	13
Results of the Phase III Experiment	16
The 49-Center Quasi-Experiment	17
The Atlanta Public School Study	19
Subsequent Analyses	20
CHAPTER TWO: SAMPLE AND METHODS	22
Selection of Sample and Sites	23
Description of Sites	26
Selection of Centers at the NDCS Sites	29
Description of the Phase III Centers	31
Description of Target Children and Families	39
Evaluation of the Sample	40
Power to Detect Effects	43
Analytic Methods and Issues	48
NDCS Approach to Multiple Regression	48
Measures of State Versus Measures of Change	51
Attrition	53
Properties of Observation-Based Behavioral Measures	54
Use of Time-Sampled Observations	56
Validity of Observations	58
Observer Effects	59
Reliabilities (Generalizabilities) of Observation Measures	61
Units of Analysis	66

TABLE OF CONTENTS
(continued)

	Page
CHAPTER THREE: THE CAREGIVER IN THE CLASSROOM	70
BACKGROUND	70
The Adult-Focus Instrument	71
Phase III Samples and Procedures	75
Introduction to the AFI Analyses	77
DESCRIPTION OF VARIABLES	78
Description of Caregiver Behavior	78
Content of Caregiver Interactions	78
Who Caregivers Interacted With	83
How Caregivers Interacted	85
The Selection and Construction of Dependent Measures	87
Reliability of the Dependent Measures	88
Regression Analyses	92
Regression Models	92
Tests for Robustness	96
Samples in the Regression Analyses	97
EFFECTS OF THE POLICY VARIABLES	100
Lead Teacher Behavior in the 49-Center Study	100
Group Composition Measures	100
Consistency of Group Size and Ratio Effect: Other Samples	113
Caregiver Qualifications	114
Consistency of Effects for Caregiver Qualifications in Other Samples	115
Covariables	116
Consistency of Effects for Covariables in Other Samples	117
Summary	118
Group Size and Ratio	118
Caregiver Qualifications	120
CHAPTER FOUR: THE CHILD--BEHAVIOR IN THE CENTER	123
BACKGROUND	123
The Child-Focus Instrument	124
Phase III Sample and Procedures	133

TABLE OF CONTENTS
(continued)

	Page
DESCRIPTION OF VARIABLES	135
Selection and Construction of Dependent Measures	135
Reflection/Innovation	138
Verbal Initiative	139
Cooperation/Compliance	139
Noninvolvement	140
Aimless Wandering	140
Task Persistence	140
Orientation to Adults	140
Orientation to Individual Children	142
Orientation to Groups	142
Other Measures	142
Reliability of the Dependent Measures	145
Approach to the CFI Analyses	146
EFFECTS OF THE POLICY VARIABLES	148
Child Behavior Results in Spring 1977	148
Covariables	148
Group Composition Variables	159
Summary of Group Size and Ratio Effects	161
Caregiver Qualifications	161
Summary of Caregiver Qualifications Effects	163
Fall/Spring Comparisons	164
Determinants of Rare but Important Events	164
Child Behavior in Structured Situations	169
CHAPTER FIVE: THE CHILD: DEVELOPMENTAL TESTS	172
Procedures and Instruments	174
The Preschool Inventory (PSI)	174
The Peabody Picture Vocabulary Test (PPVT)	176
SRI Fine and Gross Motor Tests	177
Pupil Observation Checklist (POCL)	179
Measurement of Change	180
Properties of Generalized Change Scores	186
Effects of General Background Characteristics	186
Effects of Family Process Variables	188
Center-to-Center Differences	190
Center-Level Results: The 57-Center Pooled Sample	193

TABLE OF CONTENTS
(continued)

	Page
PSI Regression Results: Overall	197
PSI Regression Results: Subsamples	201
PPVT Regression Results	209
Class-Level Analyses: The Atlanta Public Schools Study	212
Conclusions	218
CHAPTER SIX: LINKS BETWEEN CLASSROOM PROCESS AND CHILD TEST SCORES	220
Methods and Analytic Issues	222
Data Sources	222
Unit of Analysis	225
Sample	226
Results of Analyses of Classroom Process and Children's Gain Scores	227
Preliminary Analyses: Graphs and Correlations	227
Regression Analyses	230
Summary and Discussion	235
CHAPTER SEVEN: SUMMARY AND CONCLUSIONS	238
REFERENCES	245

GLOSSARY

This glossary is intended as an aid to the reader. It is not an exhaustive dictionary of terminology relevant to the study or practice of day care, but rather a list of terms used throughout the volume which may be unfamiliar to the reader or which have special meanings for the purposes of the National Day Care Study.

An alphabetical list of terms enables the reader to find any item easily; numbers refer to the location of the term in the glossary itself, which is arranged by subject area to facilitate understanding of terms in relation to each other and in the context of this study. Subject areas are:

Classification of Day Care Services
Children and Staff
Classification of Day Care Centers
NDCS Independent Variables
NDCS Dependent Variables
Statistical Terminology

Alphabetical List of Terms

activity subgroup [42]
aide [17]
auspices [21, 25]
background variable [46]
caregiver [13]
caregiver/child ratio [44]
caregiver qualifications [45]
child outcome [51]
classroom composition [38]
classroom process [49]
core care [8]
correlation [59]
cost variables [54]
day care [1]
day care center [2]
dependent variable [47]
developmental outcomes [52]
effects [48]
family day care home [3]
FFP center [34]
full-time day care [6]
funding source [30,33]
generalizability of a measure [57]
generalizability of a sample [58]
group center [23]
group day care home [4]
independent center [22,26]
independent variable [36]
infant [12]
in-home day care [5]
lead caregiver [16]
lead teacher [15]
legal status [19]
multiple regression [61]

non-FFP center [35]
nonprofit center [24]
number of caregivers [39]
outcome [53]
parent-fee
part-time day care [7]
policy variable [37]
preschooler [10]
principal components analysis [62]
private center [28]
process [50]
profit center [20]
provider [18]
public center [29]
publicly funded center [32]
regression [60]
reliability [56]
sponsored center [27]
staff [14]
staff/child ratio [43]
staffing pattern [40]
supplemental services [9]
toddler [11]
validity [55]

Classification of Day Care Services

Day Care [1] is defined as care provided to a child by a person or persons outside the child's immediate family, either inside or outside the child's home.

- A day care center [2] is defined as a licensed facility in which care is provided to 13 or more children under the age of 13, generally for up to 12 hours each day, five or more days each week, on a year-round basis.

- The term family day care home [3] refers to a private family home, generally not licensed, in which children receive care, usually for up to 12 hours each day, five or more days each week, on a year-round basis. Most state licensing codes limit family day care homes to a maximum of six children.

- A group day care home [4] is defined as a private home serving 7 to 13 children, with one or two adults.

- In-home day care [5] is defined as care provided to a child in the child's own home by a nonrelative or by a relative who is not a member of the child's immediate family.

Day care of any of these types may be either full-time or part-time.

- Full-time day care [6] is defined as care for 30 or more hours per week.

- Part-time day care [7] is defined as care for less than 30 hours per week.

The services provided by a day care center may be classified into two blocks.

- Core care [8] refers to the common components of the daily experience of all children in day care centers. Core care includes provision of meals, snacks, space and educational/play materials, arrangements for minimum health care, and various caregiver services necessary to the nurturance of young children.

- Supplemental services [9] are those services to children and their families provided by a day care center in addition to core care. For children, such services include transportation, diagnostic testing and referrals. For parents, examples are social, welfare and employment services, and parent involvement in advisory and decisionmaking capacities. Supplemental services often address fundamental needs; the term "supplemental" merely reflects the fact that they are outside the scope of a minimal center day care program.

Children and Staff

The following terms are applied to children and adults in day care settings.

- Preschoolers [10] are defined as children three, four and five years of age (36-71 months). In some states most five-year-olds attend kindergarten and thus are considered school-aged children. In these cases, preschoolers are predominantly 36 through 59 months of age.

- Toddlers [11] are defined as children aged 18 through 35 months of age.

- Infants [12] are defined as children from birth through 17 months of age.

- A caregiver [13] is a person who provides direct care to children in a day care center classroom, a family day care home, or in a child's own home. Unless otherwise specified, the terms caregiver and staff [14] are interchangeable in NDCS documents.

- A lead teacher [15] (or lead caregiver [16]) is the principally responsible caregiver in a day care classroom. The term "teacher" is not intended to connote a school-like atmosphere in the day care center. The term caregiver has been used to refer to persons working with children in day care settings, and the term lead teacher is sometimes used to distinguish the principally responsible caregiver in a day care classroom from her aides.

- An aide [17] is a caregiver who assists a lead teacher in a day care classroom.

- A day care provider [18] is a person who is directly or indirectly involved in the provision of day care services; including caregivers, center directors and owners.

Classification of Day Care Centers

Day care centers are classified according to legal status [19] as profit or nonprofit.

- Profit centers [20] are further classified according to auspices [21] as independent centers or group centers.

  --Independent centers [22] are not part of a chain of day care centers.

  --Group centers [23] belong to a chain (group) of day care centers.

- Nonprofit centers [24] are classified according to auspices [25] as independent centers or sponsored centers.

  --Independent centers [26] are not sponsored by any group or agency.

  --Sponsored centers [27] are classified as either private or public according to the nature of the sponsoring agency.

    --Private centers [28] are sponsored by a private agency, such as a church. (Note that all profitmaking centers, as well as independent nonprofit centers, are necessarily private.)

    --Public centers [29] are sponsored by some government agency, such as a city school system or a county welfare department.

In addition to classification by legal status and auspices, day care centers may be classified by a cross-cutting typology according to funding source. [30]

- Parent-fee centers [31] derive more than half of their income from parent fees.

- Publicly funded centers [32] derive their funding principally from government subsidies and gifts and contributions.

Alternatively, centers may be classified by funding source [33] according to federal financial participation (FFP). This typology was used in Supply Study analyses, and the reader may find these terms used when Supply Study data are referred to.

- An FFP center [34] is defined as any center which serves one or more federally subsidized child(ren).

- A non-FFP center [35] is defined as a center which serves no federally subsidized children.

NDCS Independent Variables

NDCS independent variables [36] are those variables whose costs and effects were to be measured. There are two types of independent variables: policy variables and background variables.

- Policy variables [37] are those characteristics of day care centers which may influence the quality and cost of center day care and which are or can be affected by federal policy. The NDCS was concerned with two major classes of policy variables: classroom composition and caregiver qualifications:

  --Classroom composition [38] describes configurations of caregivers and children in day care classrooms. Classroom composition is defined by three variables. (Note that any two of these variables mathematically define the third.)

  --Number of caregivers [39] is defined as the total number of caregivers assigned to each classroom. (The term staffing pattern [40] may refer not only to the number of caregivers assigned to a classroom, but also to the mix of teachers and aides or to the mix of qualifications of the caregivers in a classroom.)

  --Group size [41] is defined as the total number of children assigned to a caregiver or team of caregivers. In most cases, groups occupied individual classrooms or well-defined physical spaces within larger rooms. In a few "open classroom" centers, children were free to move from group to group. In such cases, clusters of children participating in common activities under the supervision of the same caregiver or team of caregivers were considered to be "groups." (The term activity subgroup [42], by contrast, refers to the actual number of children interacting with a particular caregiver. A group of 20 children, for instance, might be divided into three activity subgroups, one with the lead teacher, and two with aides.)

--Staff/child ratio [43] is defined as number of caregivers divided by group size. Higher, or more stringent, staff/child ratios are those with a smaller number of children per adult. For instance, a ratio of 1:5 is higher, or more stringent, that a ratio of 1:10 (which is lower, or less stringent). Note that the terms staff/child ratio and caregiver/child ratio [44] are interchangeable in NDCS discussions.

--Caregiver qualifications [45] variables were developed to describe caregivers' years of formal education, amount of training and/or education related to child development, and amount of work experience as a caregiver.

- Background variables [46] are characteristics of day care centers which can be influenced by government regulation only indirectly, if at all. Examples are age, sex and race of children, or socio-economic characteristics of families and of the community served by a center.

NDCS Dependent Variables

NDCS dependent variables [47] are those features of day care costs and quality measured as indicators of the effects of such center characteristics as group size, staff/child ratio and caregiver qualifications (the study's independent variables).

- In NDCS discussions, the term effects [48] is often used to distinguish dependent variables pertaining to quality in day care from dependent variables pertaining to day care costs. There are two major classes of effects variables.

  --The term classroom process [49] (or process [50]) refers to the behavior of children and caregivers in the classroom; that is, the dynamics of their interaction. Process was recorded using two observation instruments, one concentrating on children's behaviors (the Child-Focus Instrument) and one concentrating on caregivers' behaviors (the Adult-Focus Instrument).

--The term child outcomes [51] (or developmental outcomes [52], or outcomes [53]) refers to children's gains in school-readiness skills; although a number of tests and ratings of social and cognitive development were field-tested, ultimately only two, both standardized cognitive tests, proved reliable enough to be used as outcome measures: the Preschool Inventory (PSI) and the Peabody Picture Vocabulary Test (PPVT).

- Cost variables [54] correspond in the main to commonly used terminology in accounting and economics. Where terms or variables peculiar to the NDCS are introduced, they are explained in the text.

Statistical Terminology

- The validity [55] of a measure is the degree to which it measures what it purports to measure. Various features of a measure may be indicative of its validity; such as: (1) a direct conceptual relationship between the measure and the construct of interest (e.g., between an observer's count of the number of children present in a class and the variable group size); or (2) agreement with other measures of the same construct (e.g., agreement between observation-based measurements of group size and schedule-based measurements of group size).

- The reliability [56] of a measure is the degree to which it gives consistent results when applied in a variety of situations; that is, the degree to which it is free of measurement error. Reliability coefficients vary from 0.00 to 1.00. A coefficient of 0.00 indicates a completely unreliable measure; a coefficient of 1.00 indicates a measure that gives perfectly consistent results across all situations. Thus, a reliability coefficient of .95 indicates that 95 percent of the measured variation among the objects of measurement (e.g., among children) is attributable to genuine differences among the objects of measurement, and that only 5 percent of the variation measured is attributable to random effects of errors of measurement.

- The generalizability of a measure [57] is a sophisticated extension of the concept of reliability in psychological measurement theory. It incorporates the notion that the numerous sources of variation in measurement groups as "measurement error" according to standard reliability theory may or may not be defined as "error," depending on one's purpose in using a given measure. [The concept of generalizability is a very complex one which cannot be clearly presented in the limited space available here. For a definitive treatment of the subject, the reader is referred to L. Cronbach, G. Gleser, H. Nanda, and N. Rajaratnam, The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles (New York: John Wiley & Sons, Inc., 1972).]

- The generalizability of a sample [58] is the degree to which the sample accurately represents a universe to which findings based on the sample are to be extended.

- The correlation [59] (degree of association) between two variables is represented by a correlation coefficient expressed as a decimal fraction. Correlation coefficients range from +1.00 (representing a perfect positive correlation) through zero (representing the absence of any correlation) to -1.00 (representing a perfect negative correlation). For example, a positive correlation between children's scores on Tests A and B would mean that children with high (or low) scores on Tests A also tend to have high (or low) scores on Test B. If the two tests' scores were negatively correlated, then high scores on Test A would tend to be associated with low scores on Test B, and vice versa.

- Regression [60] analysis is a technique for extracting from data an idealized representation, in the form of a straight line, of the relationship between two variables. That is, regression defines the particular straight line which is the "best" linear approximation of the less clearcut pattern exhibited in the data. Similarly, multiple regression [61] analysis extracts an idealized representation of the relationships between a given dependent variable and two or more independent variables.

- Principal components analysis [62] produced alternative weighted combinations of variables ("principal components"), thus allowing the researcher to select a small number of components which convey most of the important information in a data set--that is, which together account for a large proportion of the variance in the data. For example, a large number of variables related to socioeconomic status might be reduced to a few components--clusters of variables which are highly correlated with one another and only weakly related to variables in other components.

FOREWORD

Providing sound research which supports social policy directions affecting the lives of children and families is unquestionably a major goal of the Administration for Children, Youth and Families. By producing a clear signal in an often times cloudy environment, we are able to fulfill this important responsibility that has been entrusted to us.

The National Day Care Study (NDCS) is an outstanding example of our meeting this responsibility. This study has been widely recognized in both public and private sectors as one of the most important social policy research investigations ever by the Department. Its information has been widely used by many people and organizations, and it already has had a major impact on the drafting of the new HHS Day Care Regulations.

The NDCS searched for day care center characteristics which can both protect children from harm as well as foster their social, emotional and cognitive development. It discovered that these outcomes are clearly attainable when groups of children are small and when caregivers receive training in child-related areas. It also found that relaxing the staff/child ratio would not adversely affect children but could lower costs substantially and thus enable more children to receive care. That these findings held up across diverse sites and with different groups of children, provided support that all children can benefit from a single set of standards.

In all, I feel that the NDCS has more than justified the tremendous energy and time that has gone into it. Through this kind of commitment to excellence in its research programs, the Administration for Children, Youth and Families

can be an instrumental force in enhancing the well-being of all children and families.

I am pleased to present the final volumes of the study--Volumes II and IV-A, B and C. Volume II is the research companion to Volume I--"Children at the Center." It provides quantitative support to the study's major findings. Volume IV is a compendium of technical papers which address study-related background issues, NDCS measures and methods and detailed results of individual outcome areas.

Jack Calhoun
Commissioner, Administration
for Children, Youth and Families

October, 1980

PREFACE

The federal government has become a major purchaser of child care, chiefly for the children of the working poor. With the growth of federal expenditures has come increased public concern about the quality and cost of care purchased with federal dollars. The National Day Care Study (NDCS) addressed this dual concern. Commissioned in 1974 by the Office of Child Development,* the study was conducted by two private research organizations--Abt Associates Inc. and SRI International. The study concluded that, by setting appropriate purchasing standards, the government could buy better care at lower cost than it currently buys, thus allowing it to serve more children within existing budgets.

Results of the study were summarized in a report published in March 1979.[1] The results were heavily cited in supporting arguments for proposed federal regulations, which were published in the Federal Register in early 1980.[2]

The present volume is one of a series supplementing the summary report.[3] It is intended to provide professionals in developmental psychology and related fields with a description of the methods and findings underlying the study's conclusions about links between regulatable characteristics of day care centers and the experiences and development of preschool children in center care.

Policy Context of the NDCS

Public concern with the quality of federally subsidized child care is embodied in the Federal Interagency

*The Office of Child Development is now the Administration for Children, Youth and Families (ACYF).

Day Care Requirements (FIDCR), established in 1968. The FIDCR are purchasing standards, which specify the types of facilities in which the government may buy care; they are distinct from licensing requirements, set by states and localities, which specify minimum conditions which must be met in order for a facility to operate at all. Designed to prevent harm and promote development of children in federally subsidized care, the FIDCR cover a wide variety of day care center characteristics, including groupings of staff and children, staff qualifications and training, suitability and safety of facilities, center governance, and provision of supplementary services to children and families.

In 1974 a modified version of the FIDCR was attached to Title XX of the Social Security Act, which provides grants to states to purchase social services and is the single most important source of federal funds for child care. Under Title XX, states are permitted to purchase care only in facilities that meet the FIDCR, and severe financial penalties are to be levied for noncompliance. The impending implementation of the FIDCR in 1975 provoked a storm of controversy, particularly over the FIDCR's strict staff/child ratio requirements, which exceed the day care center licensing requirements of almost all states.* Critics pointed out that implementation of the ratio requirements would have severe cost consequences for providers, states and the federal government. As a result, Congress suspended implementation of the ratio requirement--although it prohibited expenditures of federal funds in centers that allowed their staff/child ratios to fall below 1975 levels--and directed the Secretary of Health, Education and Welfare (HEW) to prepare a report on the appropriateness of the

*The Title XX FIDCR require ratios of one adult to four children for ages six weeks to three years, 1:5 for three-year-olds in groups no larger than 15, and 1:7 for four-year-olds in groups no larger than 20. On average, the states allow ratios of 1:11.4 for three-year-olds and 1:13.7 for four-year-olds.[4]

Title XX FIDCR. That report, issued in 1978, concluded that federal regulation was an appropriate means of maintaining quality in subsidized care but that the existing FIDCR were in need of revision.5

The Office of Child Development (now ACYF) had initiated the NDCS before the controversy over the Title XX FIDCR erupted. The NDCS and the Appropriateness Report were entirely independent efforts. Nevertheless the authors of the Appropriateness Report made heavy use of early results from the study, incorporating a preliminary report of NDCS findings 6 as an appendix to their own report. Subsequently, NDCS staff and the government project director were consulted during the drafting of revised regulations, which began within ACYF and was completed by the Office of HEW's General Counsel. The influence of the study is clearly visible in the proposed new standards regarding caregiver qualifications and group composition (group size and staff/child ratio). While the proposed standards deviate from the specific numerical recommendations regarding ratio and group size that appeared in the NDCS 1979 summary report, basic principles are retained--notably joint regulation of ratio and group size, with increased emphasis on the latter--as are many detailed suggestions regarding methods of monitoring and enforcement.

NDCS Approach and Findings: An Overview

The 1968 FIDCR were based on the advice of practitioners and experts in fields related to child care, as well as the best research evidence available at the time. However, in 1968 there existed only limited empirical evidence to support the basic but tacit assumptions that link various provisions of the regulations to quality of care--for example, the assumption that maintaining high staff/child ratios (few children per caregiver) will increase the

quantity and quality of adult-child interaction. Nor were there data to support the assumption that regulatory control over such center characteristics as staff/child ratio, group size and staff qualifications would produce similar outcomes for children across the regions, states, sponsoring agencies and socioeconomic groups affected by federal legislation. Similarly, though a good deal was known about the different components of cost in day care, no specific evidence existed to link costs to regulated center characteristics or to quality. The NDCS attempted to fill these gaps in knowledge by identifying costs and effects associated with variations in center characteristics that were regulated or could potentially be regulated by the federal government.

The study's sponsors and designers recognized that national policymakers have many different views of the goals of day care. For example, federally subsidized day care can be seen primarily as an institution designed to free parents to work or as a source of employment for welfare recipients. However, ACYF has long been committed to the view that day care can and should foster the development of children. Hence the study focused on the quality of care from the point of view of the child--i.e., on the nature of the child's experience in day care and on the developmental effects of that experience, as measured by naturalistic observations and standardized tests. While many potentially regulatable center characteristics were examined, primary attention focused on those characteristics which seemed most central to existing regulations and most likely to affect the daily experience of the child, namely staff/child ratio, group size and staff qualifications.

Perhaps the most general and important finding of the study was that variations in regulatable center characteristics do make a difference in the well-being of children. In contrast to many earlier studies of the effects of

variations in curriculum or resource outlay in education, the NDCS showed clearly that it matters how day care classes are arranged and who staffs them. To be sure, much of what goes on in day care is not influenced by regulatable center characteristics. There is a great deal of variability in the quality of human interaction in day care settings even when the composition of the classroom and the qualifications of caregivers are fixed. Nevertheless regulatable characteristics show relationships to meaures of children's experience and of developmental change that are significant both statistically and substantively.

More specifically, for preschool children (ages 3-5), the smaller the group in which children are placed, the more they tend to engage in creative, verbal/intellectual and cooperative activity. Also, children in small groups make more rapid gains on certain standardized tests than do their peers in larger groups. When groups are larger, individual children tend to "get lost," i.e., to wander aimlessly and to be uninvolved in the ongoing activity of the group. These findings hold even when staff/child ratios are relatively high (i.e., when there are few children per caregiver).* Adding adults (usually teachers' aides) to a large group of children improves the adult/child ratio but does not necessarily result in increased engagement on the part of the child, nor improved test score gains. Significantly, children do not appear to experience more one-to-one interaction with adults when ratios are high than when they are low.

---

*In day care classrooms, unlike many public school classrooms, it is not usual to find a single adult in charge. Configurations of two or three caregivers, usually a teacher plus aides, are more common. Both the number of children and the number of adults varies significantly from classroom to classroom. It is for this reason that staff/child ratio and group size can vary more or less independently and must be examined separately. It cannot simply be assumed that large classes will have low ratios nor that small classes will have high ratios.

The behavior of caregivers toward children is also related to group or class size, but it is related to the staff/child ratio as well. In small classes and/or classes with high ratios (few children per caregiver), staff tend to devote their attention to small clusters of 2-7 children, rather than to large clusters of 13 or more. Staff in such classes also spend less time observing children passively than do caregivers in large classes and/or classes with low ratios. In addition, the staff/child ratio shows some relationships to caregiver behavior that are not found for group size. High ratios appear to make management of children easier. Also, in high-ratio classes adults spend more time with other adults and in activities not involving children, such as performance of routine chores. This outcome may suggest that high ratios benefit caregivers by providing contact with other adults and time to do necessary tasks, but it also suggests one reason why high ratios do not appear to affect the amount of one-to-one interaction between caregivers and children: in high-ratio classes some of the time potentially available for children is diverted to activities in which children are not directly involved.

On balance, NDCS findings suggest that the importance of group size as a regulatory device for influencing quality in child care may have been underestimated and the importance of staff/child ratio somewhat overestimated. This conclusion, of course, is not an argument for abandoning regulation of staff/child ratio. Not only did ratio show some positive effects, but the range of ratios examined in the NDCS was relatively narrow and relatively high. (Most centers in the study maintained classes with five to nine children per caregiver.) This range was chosen to illustrate effects of variations in ratio between levels required by the FIDCR and levels permitted by most states. Consequently, generalization of the findings to levels outside the range

established by current regulatory variations is unwarranted. Moreover, a subsidiary study of center care for children under three suggested that ratio was as important as group size in influencing quality of care for infants and toddlers. Thus, while the findings suggest that controlling ratio alone is not an effective regulatory strategy, they also suggest that ratio should be included with group size in regulations governing classroom composition.

In addition to the above findings on group composition, the NDCS showed that qualifications of caregivers also affect quality of care. While years of formal education, degrees attained and years of experience per se made no discernible difference in quality of care, those caregivers who had education or training specifically related to young children (e.g., in early childhood education, day care, special education or child psychology) provided more social and intellectual stimulation to children in their care than did other caregivers, and the children scored higher on standardized tests.

To arrive at policy recommendations, these findings were integrated with results from other components of the study which were concerned with the costs associated with the various regulatable center characteristics and with prevailing practices in staffing and group composition among centers nationally. The costs of maintaining small groups and of employing staff trained or educated in child-related fields were found to be small, whereas the costs associated with maintaining high staff/child ratios were significant. Consequently it was recommended that, for preschoolers, the group size standards of the existing FIDCR be maintained or made more stringent, while the ratio requirements be relaxed slightly. The expected result would be an improvement in the quality of care for preschoolers together with a

reduction in costs relative to those that would prevail if the Title XX FIDCR were enforced. Implementation of the NDCS recommendations would not require major disruption of current practice, since a high proportion of centers nationally already maintain both relatively small groups and staff/child ratios that are only a little less stringent than those mandated by the FIDCR,* despite claims of some providers and state Title XX administrators that the FIDCR ratios are unrealistically strict.[7] For infants and toddlers, institution of a group size standard and maintenance of the current ratio standard were recommended. It was also recommended that training or education in a child-related field be required of all individuals providing direct care to children, and that states be required to make such training available.

Purposes and Organization of this Volume

The summary report of NDCS findings, Children at the Center, focused equally on quality and cost, for a balance between the two factors was essential in addressing the concerns of the study's many audiences and in drawing useful policy conclusions. This companion volume has a somewhat different aim and is consequently more analytic than synthetic in approach. The volume is intended to give researchers and social scientists--and lay readers who are willing to struggle with some unfamiliar concepts--enough information to judge the soundness of the evidence underlying the study's conclusions about relationships between regulatable center characteristics and the outcomes of care for the child. It makes free use of the technical apparatus

*Staff/child ratios nationwide, averaging over all classes and ages of children, are 1:6.8, compared to 1:6.3 required by the FIDCR, and 1:12.5 permitted by state licensing requirements.[8]

of developmental psychology and statistics; the lay reader will find some explanation of terms in the glossary. Children at the Center.

In order to allow this volume to be read alone, without the necessity of constant cross-reference to Children at the Center, certain sections of that volume have been included here. In particular, the sections of Chapter One of this volume that address the study design and variables have been taken substantially from Children at the Center, as has the portion of Chapter Two that describes the study sample. Other sections of Chapters One and Two are new, including a fairly detailed discussion of general analytic issues and approaches. Chapters Three through Six describe instruments, analyses and results linking regulatable center characteristics to caregiver behavior, child behavior and child test scores. These chapters constitute detailed support for Chapters Five and Six of Children at the Center, which summarized the study's results on quality of care.

NDCS conclusions about the impact of different day care classroom arrangements on the child rest on convergence of evidence from several sources, rather than on any single measure or small set of measures. Relevant bits of evidence must necessarily emerge piecemeal in the chapters that follow, if procedures and findings are to be described in enough detail to convince a potentially critical audience of their adequacy and correctness. The effect on the reader may be rather like viewing a pointillist painting, first from across the room, then up close. From a distance, as in this Preface or in Chapter Six of Children at the Center, outlines are clear and a coherent picture appears. Up close, tiny points of data take on a life of their own; their relationship to the whole becomes obscure, and many points seem not to fit at all. Nevertheless, immersion in

particulars is required if this report is to serve its purpose of drawing broad outlines where the authors think they fit best, while giving readers sufficient information to draw outlines of their own.

ACKNOWLEDGEMENTS

Since the National Day Care Study began in 1974, a great number of people have participated in the effort. These include project staff at Abt Associates Inc., site staff in Atlanta, Detroit and Seattle, and a panel of consultants from across the country, all of whom were ably directed by Dr. Richard Ruopp, Project Director. Staff and consultants at the Administration for Children, Youth and Families, and in particular Mr. Allen Smith, the Government Project Officer, also provided valuable direction for the study. Individual staff and roles are acknowledged in greater detail in Volume I, Children at the Center.

Preparation of this supporting volume has been one of the study's final tasks, and one of its most demanding and rewarding. My co-authors, Dr. Barabara Goodson, Judith Singer and Dr. David Connell sifted through the results of years of their work to arrive at the summaries presented in Chapters Three, Four and Six. All of us drew heavily on the prior work and continued help of the rest of the study's analytic staff, especially Dr. Robert Goodrich, Research Director, Dennis Affholter and Dr. William Bache. Special thanks are due to Dr. Nancy Goodrich, who was not only one of the study's principal analysts, but took on the task of managing the final phase of the study--a task that she discharged with efficiency and good humor.

Producing this volume required a considerable effort by Christine Bornas, former study secretary, and Karen Hudson, secretary for this final phase of the study. They managed to prepare drafts, organize changes, make corrections and produce the final papers, always within the time schedules provided. To them, and to all of the staff, I give my warmest thanks.

Jeffrey Travers
Associate Project Director
Abt Associates Inc.
Cambridge, Mass.
October, 1980

CHAPTER ONE: INTRODUCTION

Study Objectives

The NDCS addressed three policy questions:

- How is the daily experience and consequent development of preschool children in day care centers affected by variations in staff/child ratio, group size, caregiver qualifications and other regulatable center characteristics?

- How is the per-child cost of center-based day care affected by variations in staff/child ratio, group size, caregiver qualifications and other regulatable center characteristics?

- How does the cost-effectiveness of center-based day care change when adjustments are made in staff/child ratio, group size, caregiver qualifications and other regulatable center characteristics?

The study focused on the largest group of children receiving federally subsidized care--preschool children (aged 3-5)--and on the day care settings in which most of these children are found--urban day care centers serving low-income families. The study also focused on program characteristics that have long been considered key determinants of quality and cost in center care--staff/child ratio, group size and giver qualifications.

Study Organization

The Administration for Children, Youth and Families funded two research organizations to conduct the NDCS: Abt Associates Inc. of Cambridge, Massachusetts, and SRI International of Menlo Park, California. Abt Associates had overall administrative and technical responsibility for the study, while SRI International, as testing contractor, was

responsible for selecting and administering measures of both day care classroom processes and children's development.

The main component of the NDCS, a Cost/Effects Study of center-based day care for preschoolers, addressed the above policy questions directly. The chapters that follow are concerned almost exclusively with the "effects" portion of that study, i.e. with the part of the study that examined links between regulatable center characteristics and the daily experiences and development of preschool children in a purposefully selected sample of centers. However, it is important to bear in mind that the research discussed in this volume was part of a larger effort that included not only a cost study, but also two substudies that provided invaluable supplementary information on characteristics of day care centers nationally and on center care for infants and toddlers. In addition, the research design and methods described here were developed during two preparatory phases which will not be described in detail but which were essential to the success of the project.

The first of the two supporting studies, the National Day Care Center Supply Study,[1] was a national telephone survey designed to collect information about enrollment, staffing, costs and other characteristics of centers. Unlike the Cost/Effects Study, the Supply Study was not limited to those centers primarily serving preschool children. Results were based on a national probability sample of over 3,100 centers, stratified by state. The data provided a profile of center-based care available nationally and by state, as well as estimates of compliance with state and federal regulations. Supply Study data also played an important role in projecting the national implications of the results of the cost-effects component of the NDCS and the potential impact of alternative regulations, funding policies and monitoring practices.

The second supporting study of the NDCS focused on center care arrangements for children under three. The Infant/Toddler Day Care Study was initiated after the Title XX FIDCR imposed staff/child ratio requirements for centers receiving federal funds to care for infants and toddlers. (The 1968 FIDCR had not established ratio standards for infant-toddler care.) This research effort was designed to provide policymakers with three kinds of data not previously available. First, centers caring for infants and toddlers were surveyed nationally to provide data about their distribution and characteristics, e.g., equipment, staff/child ratios, group sizes, program schedules and activities. Second, on-site interviews were conducted with selected center directors, caregivers and parents to gather more detailed data on these center characteristics, as well as opinions about infant and toddler care. Third, selected staff were observed as they cared for infants and toddlers in order to develop a profile of caregiver behavior. Caregiver behavior was examined in relation to staff/child ratio, group size and caregiver qualifications.[2]

The NDCS Cost/Effects Study was conducted in three phases. Phase I (July 1974 to September 1975) was devoted to refinement of the study design, to selection of sites and centers and to initial selection and field testing of study instruments. Atlanta, Detroit and Seattle were chosen as the study sites, and a total of 64 centers were subsequently selected for participation in Phase II.[3] Phase II (September 1975 to September 1976) was a year-long study of naturally existing relationships between regulatable center characteristics and outcomes for children. The 64 centers were selected for high or low values of staff/child ratio, group size and staff education. Measures of classroom process, based on observations of caregivers and children, and measures of developmental change, based on standardized tests and rating scales, were administered in all 64 centers. Data were analyzed to (1) formulate initial hypotheses about

relationships among regulatable center characteristics, classroom process and developmental outcomes; and (2) refine the measures of regulatable characteristics, classroom process and developmental outcomes to be used in in Phase III.[4]

Phase III (October 1976 to September 1977) was designed to answer the study's three major policy questions. The Phase III investigation had two components: a 49-center quasi-experiment conducted in all three sites, and a randomized experiment conducted in eight centers operated by the Atlanta Public Schools (APS). (The eight APS centers were not included in the 49-center sample.) In both studies, selected center characteristics were altered systematically, permitting measurement of the costs and effects associated with such changes.

Phase III Design

The quasi-experiment was a comparison of three groups of centers (Figure 1.1). Group I (the "treatment" group) consisted of 14 centers which had low observed staff/child ratios (1:9.1) in Phase II, and whose ratios were increased to 1:5.9 in Phase III.* Effects of this treatment on caregivers and children were compared with results from a matched group of 14 untreated low-ratio (1:9.1) centers (Group II) and with those from a group of 21 untreated high ratio (1:5.9) centers (Group III). The three sets of ratios applied to classrooms that served primarily three- and four-year old children. In some centers, three-year-olds were clearly separate from four-year-olds; in others, the two ages were mixed in the same classroom. No attempt was

*Note that, in conformance with HEW directives, manipulations consisted only of making low ratios higher. The Group I treatment simulates one potential effect of full enforcement of FIDCR under Title XX--namely an increase in ratios in centers serving publicly funded children but currently operating below FIDCR ratios.

Figure 1.1

DESIGN OF THE 49-CENTER QUASI-EXPERIMENT

Group I	–	Treated centers
	–	(Observed mean ratio for 14 centers = 1:9.1 in Phase II; ratio raised to 1:5.9 in Phase III)
Group II	–	Untreated low-ratio centers
	–	(Observed mean ratio for 14 centers = 1:9.1)
Group III	–	Untreated high-ratio centers
	–	(Observed mean ratio for 21 centers = 1:5.9)

made in the quasi-experiment to alter natural variations in age-grouping. Group size, caregiver experience and years of education were distributed as evenly as possible across the three experimental groups, so that the effect of ratio could be singled out. Ratio was chosen for manipulation because of its critical policy relevance; manipulation would reduce any confounding between ratio and other center characteristics, permitting relatively clearcut assessment of its effects.

The APS Study was an eight-center, 29-classroom experiment in which children were randomly assigned, within centers, to classrooms that differed systematically in level of staff education and staff/child ratio (Figure 1.2). Group size and caregiver experience were distributed as evenly as possible across the three experimental groups. Twelve of the experimental classrooms served three-year old children and 17 served four-year olds. This design made it possible to measure the main effects and interactions of staff education and staff/child ratio for children of different ages (three- and four-year olds).

Staff in the APS centers fell into three distinct categories of educational background. First, center directors (who were required to work in classrooms as well as to function as directors) had bachelor's degrees; most also had master's degrees. Second, lead teachers were graduates of the Atlanta Area Technical School (AAT) two-year post-secondary training program in day care or had completed at least two years of college. Third, aides generally had high school diplomas (or an equivalent such as the G.E.D.); the majority of aides had also completed the 60-hour state--required training courses in day care offered through AAT. As shown in Figure 1.2, persons at these three levels of education were assigned to be lead teachers in the experimental APS classrooms--some in classes with relatively high

Figure 1.2

DESIGN OF THE ATLANTA PUBLIC SCHOOLS (APS) EIGHT-CENTER EXPERIMENT

	High Ratio (Observed Mean Ratio = 1:5.4)	Low Ratio (Observed Mean Ratio = 1:7.4)
High Staff Education	4 Classrooms	4 Classrooms
Medium Staff Education	7 Classrooms	4 Classrooms
Low Staff Education	6 Classrooms	4 Classrooms

High staff education: lead teacher was a center director, usually with a master's degree

Medium staff education: lead teacher was a graduate of Atlanta Area Technical School's two-year day care program

Low staff education: lead teacher had not completed the Atlanta Area Technical School's two-year day care program

staff/child ratios, others in classes with lower ratios. Thus, ratio and education were crossed in a two-way factorial design. Children were then randomly assigned within centers to these experimentally organized classes. Random assignment, together with the fact that the children served by APS centers were unusually homogeneous in ethnic and socio-economic background (virtually all were black children from low-income families) minimized any confounding of center characteristics and children's background characteristics.

The two Phase III components addressed similar questions but had designs with different experimental strengths and weaknesses. Because the 49-center study included a large and diverse group of centers in three different sites, its results, if uniform across the sample, were likely to be widely generalizable; however, the diversity of the sample also posed challenges for analysis and interpretation. The APS study provided a greater degree of experimental control and afforded more safeguards against confounding of center characteristics with characteristics of the children, families or communities served. However, the generalizability of its results was potentially limited by the homogeneity of the sample. The relatively consistent results actually obtained from the two study components constitute a far sounder basis for policy conclusions than would findings from either component alone.

Variables and Measures

Choice of independent and dependent variables was motivated by a basic value decision made at the outset of the study by ACYF and concurred in by its contractors, namely the decision to focus attention on those aspects of the quality of day care that bear directly on the child. In effect ACYF and its contractors took the position that the primary goal of day care purchasing standards is to ensure

the best possible environment for the most children. Other goals of day care--e.g., freeing parents to work, serving as a vehicle for delivery of social services to parents, employing low-income people as staff and fostering their development as professionals--were recognized as legitimate and important but were not central to the study.

As a consequence, in selecting regulatable center characteristics for intensive investigation as independent variables, priority was given to those characteristics deemed most likely to affect children's daily experiences, namely the composition of the classroom (principally group size and staff/child ratio) and the qualifications of caregivers (education, experience and training). Other center characteristics (space, equipment and materials; center philosophy and curriculum; director qualifications; stability of caregiver/child relationships; availability of nutrition and health services; availability of other supplementary services and specialists; opportunities for parent involvement) were examined in descriptive and exploratory fashion to determine whether any appeared to have major effects on classroom processes and child outcomes.[5] However, in light of preliminary results which suggested that most of these variables had minimal effects on the particular outcome measures chosen, only a few of the variables were investigated further, and then only to a limited extent.

In selecting dependent variables and measures, priority was given to descriptors of the immediate experience and consequent development of the child. Ancillary data were collected, largely through interviews with parents and staff, on parental satisfaction, parental income and employment, delivery of supplementary services to families, staff satisfaction and professional development. Again, descriptive and exploratory analyses were conducted,[6] but

these data did not play a central role in the study's policy conclusions. Throughout the remainder of this volume, discussion focuses almost exclusively on the study's major independent and dependent variables.* Other variables are treated briefly in Children at the Center and in a volume of technical appendices.

Independent Variables and Measures

Independent variables were of two types: background variables, such as age, sex and race of children, and socioeconomic characteristics of families and of the community served by the particular center, and policy variables, i.e., center characteristics subject to regulatory control. While background variables are unregulatable and therefore not of direct policy relevance, their effects had to be taken into account in assessing the effects of the policy variables. Distributions of policy and background variables are presented in Chapter Two of this report.

Background Variables

Information on background characteristics of children and their families was gathered through interviews with parents. Background information included family income, sources of income, parents' education and occupation, length of parents' employment, number of siblings and number of adults living in the house. Age, sex and race of children were verified. In addition, census data were used to provide background information on demographic characteristics of the community, chiefly its socioeconomic and racial composition.

---

*Some of the secondary data are used in Chapter Five in exploring factors related to children's test performance.

Policy Variables: Definitions

As indicated earlier, the major policy variables examined in the NDCS fell into two categories--those relating to classroom composition and those relating to caregiver qualifications. Three variables fell under the rubric of classroom composition:

- number of caregivers, defined as the total number of caregivers present in or assigned to a classroom or group of children;*
- group size, defined as the total number of children present in or assigned to a class or to a principally responsible caregiver;* and
- staff/child ratio, defined as the number of caregivers divided by group size.

Caregiver qualifications variables included of total years of formal education, presence or absence of education or training specifically related to young children, and day care experience (both years of experience prior to current job and duration of employment in current center).

Policy Variables: Measures

Information on variables related to classroom composition was gathered by two methods, one based on

*In all but a few NDCS centers, groups of children were assigned to particular rooms, supervised by a single caregiver or several caregivers. In a few "open classroom" centers, however, very large numbers of children (approaching 100 in extreme cases) were present in a single large room. Even in such centers, children clustered around individual caregivers or small teams dispersed around the room, though children were often free to move from group to group. Numbers of children in these smaller groups constituted the group size used for NDCS analytic purposes. Similarly, numbers of caregivers were the number of adults in physically separated groups.

schedule or roster data and the other on direct observation. Schedule-based and observation-based measures of classroom composition were not always in close agreement. Differences between the two were primarily attributable to two phenomena--absenteeism and merging of classes. Because observations capture the group configurations actually experienced by the child and because they automatically take account of absenteeism and merging, observation-based measures were used in all the analyses reported in this volume. However, because of the importance of these issues for monitoring and enforcement, comparative investigations of the two types of measures were conducted and are reported elsewhere.7

Three sets of observation-based data on classroom composition were collected. One set of counts was made in conjunction with behavioral observations of caregivers, and a second in conjunction with observations of children; these counts were used in the corresponding behavioral analyses. (Behavioral observations are described below and in later chapters.) A third set was collected on a regular basis by NDCS staff employed full time at each center during Phases II and III; this set was used in analyses of children's gains on standardized tests, which were expected to reflect classroom configurations prevailing over the year, rather than at any particular point in time.

Information on caregiver qualifications was initially gathered through interviews with nearly all caregivers who worked in the study's "target" classrooms--those serving primarily three- and four-year old children. In analyses of the relationship between caregiver qualifications and caregiver behavior, which used the individual caregiver--teacher or aide--as the unit of analysis, the qualifications of the individual in question were used directly as independent variables. In analyses of effects on child behavior, qualifications of teachers and aides within each classroom

were averaged together, and classes were the units of analysis. In analyses of effects on children's test scores, qualifications of lead teachers (not aides) were averaged to center level, and centers were the units of analysis. (Reasons for these choices of units of analysis are given in Chapter Two.)

Dependent Variables and Measures

Choosing dependent variables and measures to capture the child's experiences in the classroom and assess consequent changes in the child's development was perhaps the most challenging conceptual and practical task facing the NDCS. At the outset of the study there existed no universally accepted catalogue of desirable experiences, traits, skills and behaviors, nor does such a catalogue exist now. And even when the desirability of some experience or outcome was widely agreed upon in principle, adequate measures often did not exist. For example there is fairly widespread agreement that an ideal care environment should build a child's self-concept, but instruments for measuring self-concept in preschoolers are still being developed by basic researchers.

After a long process of experimentation and adjustment, chronicled in reports issued at the ends of Phase I and Phase II,[8] an empirical strategy of measurement and analysis evolved. The strategy relied heavily on two observation instruments selected by SRI in Phase I. The two instruments, one focused on caregivers and one on children, use trained, on-site observers to record everyday classroom behavior in considerable detail. From the resulting records of frequencies of specific behaviors, measures of broader variables were constructed, usually by summing frequencies of behaviors that were conceptually related and

empirically correlated.* For example, a caregiver behavior variable called "management" was constructed by summing the frequencies of the behaviors "commands" and "corrects," which are recorded directly. In addition, two standardized tests, designed to measure selected school-related cognitive and linguistic skills, were administered to each child. In short, the study attempted to describe as objectively and comprehensively as possible the behaviors associated with various configurations of regulated center characteristics, and to supplement this information with information about children's test performance. The study's conclusions and policy recommendations rest on largely post hoc value judgments about the total pattern of caregiver behavior, child behavior and test scores found to be associated with the different regulatory variables.

The observation instruments, tests and variables constructed from them are described in detail in Chapters Three through Five. At this point, variables are simply listed with a brief, general explanation for each of the three broad domains:

Caregiver Behavior. Variables in the domain of caregiver behavior primarily characterize the nature and number of contacts between caregivers and children. The variables distinguish warm, stimulating child-directed behavior from more passive and instrumental forms of behavior. They also distinguish interaction directed at individual children and small groups from interaction directed at larger groups and other adults. Variables in this domain include:

---

* In a few cases, frequencies of individual behaviors were treated as variables directly and in other cases methods of combination other than simple summing were employed. Details are provided in later chapters.

- Social Interaction with children (praising, comforting, responding, questioning and instructing);
- Management of children (commanding and correcting);
- Observing children;
- Center-Related Activities (planning, arranging materials, cleanup, recordkeeping, etc)
- Overall frequencies of all types of interaction with
  --individual children
  --small groups (2-7 children)
  --medium groups (8-12 children)
  --large groups (13 or more children)
  --other adults

Child Behavior. Variables in the domain of child behavior characterize both the child's social interactions and solitary activities, as well as relative amounts of interaction with adults, other children and objects in the physical environment. The variables distinguish activities of a verbal/intellectual and/or social nature from behavior indicating passivity or withdrawal. Variables in this domain include:

- Verbal Initiative (giving opinions, preferences, information or comments);
- Reflection/Innovation (considering, contemplating, tinkering, or adding a new idea or new object to an ongoing activity);
- Cooperation/Compliance (active, appropriate responding to questions, requests, and commands from adults and other children);
- General Interest/Participation in center activities;
- Aimless Wandering;
- Noninvolvement in tasks or activities;
- Task Persistence (duration of longest activity in an observation period);
- Attention to Adults;
- Attention to Other Children;
- Attention to the Environment.

Test Scores. Variables in this domain were gains from fall to spring on two standardized tests:

- The Preschool Inventory, a global test of school-related skills and knowledge, including knowledge of shapes, sizes, parts of the body, spatial relationships, etc.

- The Peabody Picture Vocabulary Test, a measure of receptive vocabulary in which the child matches words and pictures.

The tests were not assumed to measure general cognitive or linguistic ability or development; moreover their cultural biases were acknowledged. They were included as outcome measures because of their potential for predicting the child's success in elementary school--a concern of many parents and providers. Fall-to-spring gains were calculated using techniques designed to circumvent certain well-known technical problems involved in measuring change. (See Chapter Five).

Results of the Phase III Experiments

Results of the Phase III experiments suggest that the regulatory variables chosen for experimental manipulation--primarily staff/child ratio and secondarily staff education--have few detectable effects on the behavior of caregivers, the behavior of children or children's test scores. High staff/child ratios did appear to have some positive effects, but these effects were neither consistent nor large and may have been due to chance. Results of the experiments are reported briefly in this introductory chapter in order to clear the way for discussion of more fruitful analyses of nonmanipulated variables, to be reported in subsequent chapters.

The 49-Center Quasi-Experiment

The question of central interest in the quasi-experiment was whether the experimentally induced increase in staff/child ratio would produce more desirable outcomes in treated centers than in the matched group of untreated, low-ratio centers. (Would Group I (treatment) differ from Group II (low-ratio comparison) in observed behavior of caregivers or children, or in children's test scores?) The comparison group of untreated, naturally high-ratio centers (Group III) was included to address a supplementary question: Would the experimental increase in ratio eliminate most or all differences between centers that previously operated at different ratios, or would differences in outcomes continue to exist, presumably because of other center characteristics that normally accompanied high ratios but were unaffected by the experimental increase in ratio? (That is, would Group III (untreated high-ratio) differ from Group I (treated high-ratio)?)

Answers to these questions were provided by a series of one-way analyses of variance, using the three groups as levels of an independent, classificatory variable and using a variety of behavioral measures, as well as test score change measures, as dependent variables. The behavioral measures included not only the constructs listed earlier but also many of the finer-grained behavioral codes from which the constructs were built. The null results were so consistent across dependent measures that it is extremely unlikely that any regrouping of codes to form new constructs would change the conclusions appreciably.

In the domain of caregiver behavior, seventeen dependent measures were examined, including all of the constructs listed earlier and all of their component codes.

Lead teachers and aides were examined separately. For lead teachers, only two codes showed significant or even marginally significant ($p<0.1$) overall differences in frequency across the three groups. The frequencies of the codes "corrects" and "responds" were lower in naturally high-ratio centers than in treatment and control centers, which did not differ from each other--a result clearly not attributable to the experimental manipulation, but to other characteristics of naturally high-ratio centers. For aides, only one marginally significant difference, potentially attributable to the ratio manipulation, appeared: aides in treated high-ratio classrooms and naturally high-ratio classrooms devoted less attention to the physical environment than did those in low-ratio classrooms.[9]

In the domain of child behavior, twenty individual codes and global constructs were examined. Separate analyses were conducted for observations made during periods of free play and those made during teacher-directed activity. For only one dependent variable was there a clear and significant ($p<.05$) effect of the ratio treatment in both types of activity periods: during both free play and teacher-directed activity aimless wandering was more frequent in low-ratio centers than in treated centers or naturally high-ratio centers. A few other significant or marginally significant overall group differences were found, but, except for the result just cited, none of the findings suggested that the experimental ratio increase had increased the frequency of desirable behavior or decreased the frequency of undesirable behavior.[10]

In the domain of test scores, no significant effects were found. Neither gains on the PSI nor gains on the PPVT differed significantly across the three groups.[11]

Considering the large number of tests performed, some of the "significant" findings alluded to above are probably due to chance. Even if taken at face value, the results do not make a persuasive case that the experimental ratio increase significantly affected either the child's social experience in the classroom or his or her development as measured by standardized tests.

The Atlanta Public School Study

With respect to the effects of staff/child ratio, results of the APS study confirmed most of the null findings of the 49-center study. In addition, the APS study suggested that formal education of the caregiver, as defined by the three levels examined in the study, had little or no effect in the classroom.[12]

As indicated earlier, the APS study had a factorial design, with two levels of staff/child ratio crossed by three levels of staff education. A series of two-way ANOVAS was performed, using as dependent variables a total of 53 measures derived from observations of caregivers and children, in addition to gain scores on the PSI and PPVT.* Of the 53 behavioral measures, ten showed significant ($p<.05$) effects due to ratio, education or their interaction. Virtually all

*APS analyses were complicated by the fact that the factorial design shown in Figure 1.2 could not be replicated in every APS center, since centers were not large enough to permit the necessary number of classes. (Three levels of education by two levels of staff/child ratio by two age groups--three- and four-year-olds--yields a twelve-celled design, ideally requiring twelve classes per center. Few centers had more than four classes.) Consequently, possible confounding effects due to center differences had to be examined before any effects could be attributed to the experimental changes induced within centers. Fortunately, exogenous center effects did not prove to be a significant confounding factor.

of the significant effects were observed in caregiver behavior rather than child behavior. Most were due to education or the interaction of education and ratio, not to ratio alone. Overall the pattern did not suggest that caregivers with more formal education provide better care for children. Instead, the pattern suggested that the APS experiment itself had introduced some anomalous behavior patterns in the classroom; for example, highly educated center directors, assigned to the role of lead teachers, continued to perform their directorial duties and consequently diverted time from interaction with children to administrative matters and hence showed more "center-related activity" than other caregivers.

Analyses of the impact of ratio, staff education and their interaction on children's gains on the PSI and PPVT were conducted separately for three- and four-year-olds, as well as for the two age groups pooled. Here one significant effect emerged: Three-year-olds made more rapid gains on the PSI in high-ratio classes. No other effects were observed.

In short, the APS study, like the 49-Center Study, showed isolated positive effects for high staff/child ratios but did not provide evidence of large or widespread effects. Caregiver education was related to caregiver behavior, but not in such a way as to suggest that more educated staff provide better care. Caregiver education showed virtually no direct positive effects on children's experience or development.

Subsequent Analyses

The essentially null results of the two experiments--if genuine and not merely due to unsuspected design flaws or lack of statistical power--would have significant

implications for regulatory policy. Therefore, to assure the validity of these results, the NDCS pursued its analyses much further. There was, within each of the various experimental groups of centers and classes, a great deal of variation not only in the experimentally manipulated variables (ratio and staff education) but also in other regulatable characteristics--group size, staff experience and child-related content of caregivers' education or training. These naturally occurring variations were examined, though multiple regression analysis, in relation to the dependent variables listed earlier. In a general sense, these analyses confirmed the experimental results already reported--that variations in staff/child ratio (within the range studied in the NDCS) have some effects, but fewer than generally believed, and that the formal education of caregivers is a relatively unimportant influence on the child's experience in day care and his or her test performance. However, other regulatable center characteristics, notably group size and education or training in fields specifically related to young children, did show important relationships to outcomes for children. Subsequent chapters describe in detail the methods and findings of these further investigations.

CHAPTER TWO: SAMPLE AND METHODS

As implied at the end of Chapter One, the analytic approach of the NDCS was essentially correlational and exploratory. In the absence of any important effects attributable to the regulatory variables which were manipulated in the two experiments, the study examined patterns of association between behavioral measures and test scores, on the one hand, and naturally varying regulatable center characteristics on the other. Natural variation included both variation in staff/child ratio and staff education within the experimental groups established in the two studies, and variation in other characteristics such as group size, staff experience and the content of staff education and training which had not been altered experimentally but had been balanced in distribution across the experimental groups.

Relationships were explored by means of multivariate statistical techniques, chiefly multiple regression. Clearly, this type of analysis does not permit firm causal inferences, although associations may suggest causal hypotheses. Nevertheless, associational findings are useful to the policymaker in setting purchasing standards for child care. Such findings identify center characteristics which are likely to be accompanied by desirable experiences and developmental outcomes for the child, even if those center characteristics do not themselves cause desirable outcomes to occur. Center characteristics that have this property can be used as benchmarks or indicators of quality in setting purchasing standards.

The success of a correlational study depends heavily on the nature of the sample, especially on the distributions of independent variables within the sample,

and on the statistical techniques used to dissect relationships between variables. This chapter sets the stage for the presentation of findings by describing the NDCS sample in some detail, focusing on distributions of independent variables, and by outlining some of the more important features of the study's statistical approach. Subsequent chapters describe dependent variables and measures in each of the three domains studied--caregiver behavior, child behavior and test scores--and present the study's main findings in each domain.

Selection of Sample and Sites

Criteria for selecting the centers to be studied in the NDCS were designed largely to maximize representation of policy-relevant centers--those serving or eligible to serve low-income children receiving subsidized care. Additional criteria were dictated by research considerations, such as cost of data collection, adequacy and stability of the sample, and feasibility of measurement. Selection criteria required that centers in the sample:

- be licensed day care centers, located in urban areas, and serving or eligible to serve federally subsidized children. Licensing is a precondition for purchase of subsidized center care. Centers were chosen over family day care homes because they supply 80 percent of licensed day care slots and receive a large portion of federal day care subsidies. Urban centers were chosen both for logistical reasons and because licensed center care is predominantly urban. The sample included both centers funded primarily by the federal government and centers funded primarily by parent fees.

- provide year-round full-day care. Only full-time year-round centers offer day care arrangements which satisfy a major intent of federal day care appropriations under Title XX--promoting parents' economic self-sufficiency by freeing them for training and work. Thus, to be eligible for participation in the study, a center had to be open at least seven hours per day, five days per week and ten months per year.

- have been in operation at least one year. To increase the probability that centers would continue in operation throughout Phases II and III, and to avoid studying non-recurring start-up behavior, centers were required to have been in operation for at least one year at the time they were selected.

- serve English-speaking preschool children. Because preschool children aged three through five constitute the majority of the day care population, they were a high priority study group. Children from non-English-speaking families were not included in the research sample for two reasons. First, adequate test batteries for non-English-speaking children did not exist. Second, non-English-speaking children consitute a small percentage of the day care center population.

- have an adequate sample of full-time three- and four-year-old children. To ensure that start-of-year and end-of-year test data would be available for an adequate number of children, centers were included in the sample only if they had 15 or more three- and four-year-old children enrolled on a full-time basis.

The study's three sites--Atlanta, Detroit and Seattle--were chosen to be as diverse as possible, in order to determine whether regulatable center characteristics have different costs or effects in different geographic, demographic

and regulatory environments.[1] Four general criteria were used for site selection. Sites had to have enough eligible centers, each with adequate distributions of the policy variables, to allow full implementation of the study design. To test for potential differences in effects due to geographic factors, the sites had to represent different geographic regions. Sites also had to differ in demographic and socio-economic characteristics in order to test for potential differences in effects associated with differences in community characteristics. Finally, sites had to exhibit regulatory diversity to test for differences in findings attributable to state and local regulatory policies.

During Phase I, socioeconomic information on 50 urban areas, obtained from census data, licensing authorities and other governmental sources, was used to identify 17 potential study sites meeting the above criteria. Most of the 33 disqualified cities were ruled out because they did not have enough eligible centers for full implementation of the study. Seven of the 17 potential sites were in the South, five were in the North and Midwest, and five were in the West.

A telephone survey of a 25 percent stratified random sample of centers in these 17 cities was conducted to determine whether centers showed distributions of staff/child ratio, group size and staff education required by the Phase II design. In addition, a further analysis of census data was undertaken in order to assure generalizability of findings. Each potential site had to be representative of a larger group of cities in the country with similar social and economic characteristics. To determine which of the 17 cities met this requirement, the entire set of 29 U.S. Census summary socioeconomic variables was used to cluster all 248 urbanized areas in the United States into a few groups.[2] Principal components analysis was employed to

compute a "measure of distance" among cities and to group them according to measures of socioeconomic status. On the basis of this analysis, together with telephone survey data, six representative cities, each of which could sustain a complete experimental design for Phase II, were chosen as potential sites:

South	North	West
Atlanta	Chicago	Los Angeles
New Orleans	Detroit	Seattle

A more intensive telephone survey, together with site visits to test the feasibility of study implementation in each of the six cities, resulted in the final choice of Atlanta, Detroit and Seattle as sites for Phases II and III.

Description of Sites

Purposeful selection of sites resulted, as intended, in demographic and regulatory diversity across sites.[3] Of the three sites, Atlanta had the highest proportion of female-headed families (12.4%) followed by Detroit (11.2%) and Seattle (9.3%). Only Seattle fell below the national average of 11 percent. Among women over 16 years of age, the highest percentage employed was in Atlanta, and this difference was even more pronounced among mothers of children under six: in Atlanta, 48.8 percent were employed; in Seattle, 29.5 percent; and in Detroit, 22.5 percent. (At the time of selection, for the U.S. as a whole, 31.1 percent of women over 16 with children under six were employed.) Atlanta residents had the lowest mean family income ($12,160), followed by Seattle ($13,233) and Detroit ($13,532). In addition, the highest percentage of families fell below the poverty line in Atlanta.[4]

The three sites also differed in regulatory climate. Although during the time of the study, state regulations in all three sites addressed issues such as space requirements, staff qualifications, safety standards and the like, Georgia's day care regulations were particularly comprehensive and detailed. In contrast, Michigan's regulations were brief and applied to nursery schools as well as day care centers; thus no regulatory distinction was made in Michigan between a preschool which cares for children for only a few hours a day and a day care center in which children are in care for a much longer period. Washington's regulations fell in the middle: Washington regulated day care centers but did not regulate nursery schools, and its day care regulations were less detailed than Georgia's.

All three states specified staff/child ratio by age of child, although none of the required ratios were as stringent as those mandated by the FIDCR. Only Georgia regulations specified maximum allowable group size according to age of child. The three states varied also in staff qualification requirements. In Georgia, both directors and classroom staff were required to show evidence of recent training in child care, although this training did not have to be in a degree program. Michigan required that the center director have a minimum of two years' study at the college level. Washington's regulations specified that program supervisors must have two years' background and experience in programs serving children and must have accumulated 45 credit hours of college or other training in child development (or have a plan to obtain such training).

Implementation of both Title XX and the FIDCR varied from site to site. At the time the sites were selected for the NDCS, Georgia required that centers serving federally subsidized children comply fully with a 1972 draft

version of the FIDCR which was never adopted by HEW. The State of Washington had established no separate system of monitoring centers specifically for compliance with the provisions of the 1968 FIDCR, relying instead on existing licensing personnel and, as elsewhere, compliance was never vigorously sought. In contrast, Michigan had initially responded to the FIDCR by seeking and receiving a limited waiver from all FIDCR provisions for some of its centers. In 1969, three levels of certification were established in Michigan--full compliance with the FIDCR, waivered certification, and noncertification. However, when Title XX was implemented and the FIDCR staff/child ratio requirement suspended, this system was dropped, and the state no longer required that centers serving subsidized children meet the FIDCR staff/child ratios, although these centers were asked to comply with the other provisions of the 1968 FIDCR.

With the advent of Title XX, Georgia decided to contract with centers for the provision of subsidized care; children eligible for such care could be sent only to centers already under contract to the state. This practice differed from that of the other two sites, where parents of children eligible for subsidized care could enroll their children in any licensed center. The center then contracted with the state for reimbursement. Thus parents of eligible children in Seattle and Detroit had a greater degree of choice in determining which center best met their individual needs than did parents living in Atlanta.

Sites also varied in the amount and type of training that was readily available. In all three sites it was possible to obtain training in day care at the college level, but only in Atlanta was training available that was designed specifically to meet minimum day care licensing standards. This training program, offered by the State

Board of Education through the Atlanta Area Technical School consisted of two basic courses in day care skills and child development. It had to be taken by all caregivers within at three years of center employment. The Atlanta Area Technical School also offered a two-year post secondary program for day care workers as well as a training course in administration for day care directors. Other day care programs in Atlanta included a graduate program for day care directors at Georgia State University and undergraduate courses in day care at Atlanta University. In addition the Georgia Department of Human Resources provided workshops run by its licensing consultants for staff in day care centers.

In Seattle, day care training was primarily provided by the community colleges. Seattle Central Community College had a two-year program of day care training, and five other community colleges offered day care courses, as did Rentnor Vocational School. The community colleges also sponsored workshops for day care staff and provided in-service training. The Puget Sound Association for the Education of Young Children, the 4-C Program (Community Coordinated Child Care) and the Seattle Child Care Resource Center also were important sources of training outside the educational institutions.

In Detroit, two-year programs were offered by Wayne State University, Wayne County Community College, Highland Park Community College, Madonna College, Mercy College, Marygrove College and Schoolcraft College. Madonna College also had a one-year program for child care aides. In addition, the Merrill-Palmer Institute trained students to work in day care centers.

Selection of Centers at the NDCS Sites

Within sites, centers were initially selected to meet the requirements of the Phase II natural study design.

The factorial design required centers with all possible combinations of high or low levels of staff/child ratio, group size and staff education--a total of eight different center types. To ensure coverage of the policy-relevant range for each policy variable, data from the Phase I telephone survey and site visits were used to select Phase II centers in which levels of regulatable characteristics varied from minimum standards set by state licensing requirements to the more stringent levels required by the FIDCR. Centers were also selected to vary as much as possible in nonregulatable characteristics. For example, efforts were made to recruit centers that operated under a variety of auspices and drew their funds from different sources, both public and private.

Diversity was also sought among the children and families served by study centers. Centers serving substantial numbers of both black and white children were included, both integrated centers and those predominantly serving children of one race. Similarly, the sample was selected to include centers serving both low- and middle-income families and therefore to include substantial numbers of children supported by public subsidy as well as children supported by parent fees.

Most Phase II centers were retained in Phase III, though some centers were dropped and others were added to meet Phase III design requirements. Nine of the 64 Phase II centers were operated by the Atlanta Public School system. Four of the latter were dropped because they did not contain enough classrooms to implement the APS design, and three larger APS centers were added in their place. Of the remaining 55 Phase II centers, six were dropped, either because they closed, declined to participate or proved to be atypical or unstable in organization during Phase II; the remaining 49 centers were retained for the quasi-experiment.

Description of the Phase III Centers*

At the beginning of Phase III, approximately 1600 three- and four-year-old children were enrolled on a full-time basis in the 57 study centers.** About 300 staff were employed as teachers or aides in the study's target classrooms--those serving primarily three- and four-year-old children.

As intended, Phase III centers showed a broad range of configurations of classroom composition.*** Across all 57 centers, observed groups sizes ranged from eight to 36, with an average of 17.6 children per group (Figure 2.1A). Most centers (75%) had group sizes between 12 and 24. Number of caregivers per classroom ranged from one to more than five, with an average of 2.4; most classes had three or fewer caregivers (Figure 2.1B). Observed staff/child ratios in target classrooms averaged 1:6.8, with a range from 1:4.2 to 1:16.4, although most centers (85%) had ratios between 1:5 and 1:9 (Figure 2.1C). Figure 2.1 also shows how the NDCS centers compare to centers nationally in distributions of the policy variables. National data are drawn from the NDCS Supply Study.5

---

*For the purposes of summary, classrooms from the 49 center study and the eight-center APS supporting study are described together in this section. Important differences between the two samples are noted where relevant.

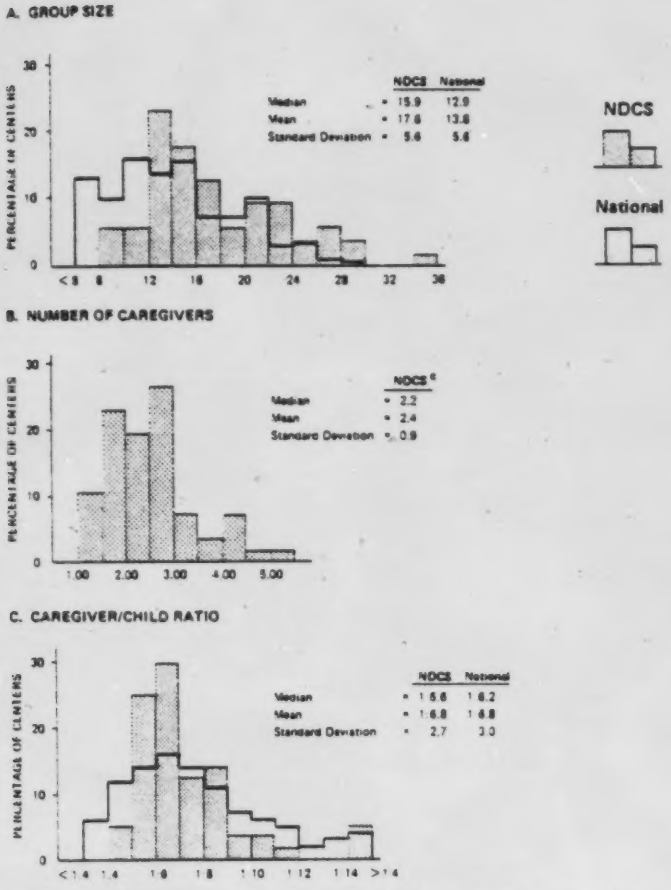
**Total enrollment in these centers was approximately 2300 children, including children under three and over four.

***A comparison of the NDCS sample and the Supply Study national sample of centers or the major policy variables is presented in the final section of this chapter. A description of other center characteristics nationally is presented in Appendix A of Children at the Center and Volume III of the NDCS Final Report. (See Preface references 1 and 3 for complete citations.)

Figure 2.1

DISTRIBUTION OF CLASSROOM COMPOSITION MEASURES (OBSERVED[a]) FOR THE NDCS AND NATIONALLY

(Center Level: NDCS N = 57; National Sample N = 3167[b])



[a] Observed group size is smaller (and therefore ratio higher) than enrolled group size and ratio, by about 12% due to child absenteeism.

[b] NDCS measures of classroom composition were taken by observation; the NDCS Supply Study gathered enrollment and planned staffing data in its survey of 17.3% of all centers nationally. These data have been adjusted for child absenteeism for the purposes of comparison.

[c] Data not available for the national sample.

Staff/child ratios were roughly comparable across sites, although Atlanta centers (both APS and non-APS) tended to have somewhat higher ratios than Detroit or Seattle centers (Table 2.1). Detroit centers tended to have appreciably larger groups than did Atlanta or Seattle centers. Seattle had the fewest caregivers per class.

Table 2.1

DISTRIBUTION OF CENTER-LEVEL AVERAGES OF CLASSROOM COMPOSITION VARIABLES

	NDCS CENTERS[a]					
	Breakdown by Site				All NDCS Centers	Centers Nationally[b]
	Atlanta		Detroit	Seattle		
	APS (N=8)	non-APS (N=20)	(N=13)	(N=16)	(N=57)	(N=3167)
Classroom Composition (Observed)						
Group Size (Number of children)	17.0	16.9	20.0	16.7	17.6	13.8
Number of Caregivers/ Classroom	2.5	2.6	2.6	2.1	2.4	c
Staff/Child Ratio	1:6.3	1:6.3	1:7.4	1:7.2	1:6.8	1:6.9

[a] NDCS policy variable data are for target three- and four-year-old classrooms averaged to the center level.

[b] Based on NDCS Supply Study data averaged to the center level. The composition variables are based on classroom data from classrooms nationally meeting NDCS target classroom criteria, and have been adjusted for absenteeism.

[c] Group-by-group data on the number of caregivers per classroom are not directly available. An approximation can be derived by multiplying group size by staff/child ratios.

The typical caregiver had completed high school and had slightly less than two years of post-secondary education (Figure 2.2A). On the average, half of the observed caregivers had received specialized training/education in child-related areas, although substantial variation

existed in this dimension (Figure 2.2B). In general, caregivers in the NDCS had less than one year's experience in other centers (Figure 2.3A); by far the largest part of caregivers' day care experience was in their current centers (Figure 2.3B).

Educational attainment was comparable across sites (Table 2.2). More marked site variations were found in the proportion of caregivers with child-related education/ training; APS centers had heavy concentrations of such caregivers. This high degree of "specialization" in the APS sample is a function of Georgia's requirement that day care workers complete state-sponsored courses in day care within three years of beginning employment, as well as the APS policy of hiring lead caregivers with associate's or bachelor's degrees in early childhood education.

Virtually all classroom staff in the 57 study centers were female. Of the caregivers actually observed in the classroom during Phase III, half were white and half were black. Their mean age was approximately 33 years, but there was considerable variation in the sample.

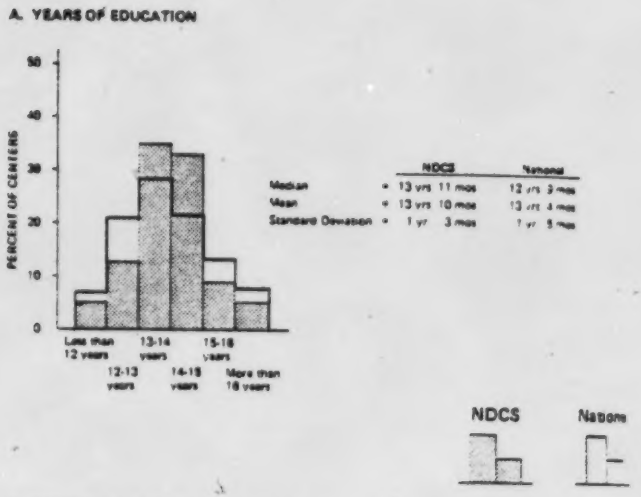
Sixteen of the 57 centers (28%) were racially integrated, where "integrated" centers are defined as those with enrollments between 20 and 80 percent black (Table 2.3). Nine centers (16%) were predominantly (more than 80%) white, and 32 centers (56%) were predominantly black.

Ten of the 57 centers (17.5%) were operated for profit, while the remaining 47 (82.5%) were nonprofit centers. Of the latter, 13 were operated by voluntary agencies, eight by public schools (the APS centers), 17 by churches, three by Head Start and six by private individuals (see Table 2.3).

Figure 2.2

DISTRIBUTION OF CAREGIVER QUALIFICATIONS MEASURES EDUCATION AND TRAINING VARIABLES FOR THE NDCS AND NATIONALLY

(Center Level: NDCS N = 57; National Sample N = 3167[b])



*Data not available for the national sample.

Figure 2.3

DISTRIBUTION OF CAREGIVER QUALIFICATIONS MEASURES EXPERIENCE VARIABLES FOR THE NDCS AND NATIONALLY

(Center Level: NDCS N = 57; National Sample N = 3167)

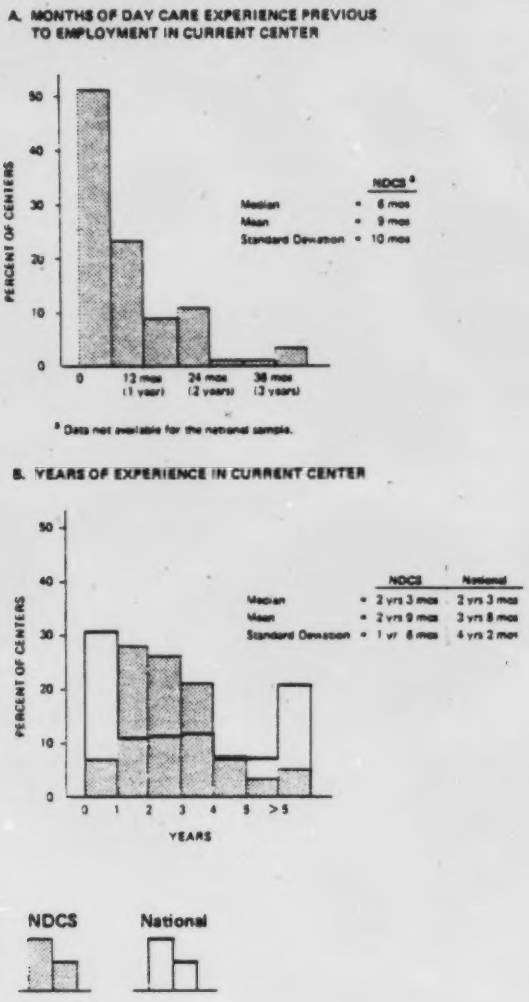


Table 2.2

DISTRIBUTION OF CENTER-LEVEL AVERAGES OF CAREGIVER QUALIFICATIONS VARIABLES

	NDCS CENTERS[a]				NDCS Centers	Centers Nationally[b]
	Breakdown by Site					
	Atlanta		Detroit	Seattle		
	APS (N=8)	non-APS (N=20)	(N=12)	(N=16)	(N=57)	(N=3167)
Caregiver Qualifications						
Years of Education	13 yrs, 8 mos	13 yrs, 5 mos	13 yrs, 11 mos	14 yrs, 6 mos	13 yrs, 10 mos	13 yrs, 4 mos
Percent of Caregivers Staff with Child-Related Education/ Training	85%	56%	29%	53%	53%	NA[c]
Previous Day Care Experience	12 mos	9 mos	6 mos	9 mos	9 mos	NA[c]
Experience in Current Center	3 yrs, 6 mos	3 yrs, 3 mos	2 yrs, 5 mos	2 yrs	2 yrs, 9 mos	3 yrs, 8 mos

a NDCS policy variable data are for target three- and four-year-old classrooms averaged to the center level.

b Based on data from all classrooms nationally, regardless of age.

c Information not available.

Table 2.3

NDCS CENTERS: SUMMARY DATA

(N = 57 Centers)

	Number of Centers	Percent of Sample
Integration		
Integrated centers (20-80% black enrollment)	16	28
Predominantly black (more than 80% black enrollment)	32	56
Predominantly white (more than 80% white enrollment)	9	16
Legal Status		
For-profit centers	10	17.5
Non-profit centers	47	82.5
Operated by voluntary agencies	13	
Operated by public schools	8	
Operated by churches	17	
Operated by Head Start programs	3	
Operated by private individuals	6	

NDCS centers tended to be located in areas of high day care demand. The typical center was in a census tract in which about half of the women were in the labor force in 1970. Approximately 18 percent of the families in census tracts surrounding study centers fell below the poverty line in the 1970 census.

Description of Target Children and Families

As indicated above, every attempt was made to achieve wide diversity in the children and families included in the study while still adequately representing children from low-income working families because of their special policy relevance. Examination of the socioeconomic characteristics of children and families in the study sample shows that these efforts were successful. Most of the background information was collected from parents at the start of Phase III and was available for most of the sample of children who were observed and tested.*

Slightly over half of the target families were single-parent households. Three-fourths of all mothers were employed either full- or part-time and the remainder were in school or a job training program. About 90 percent of the fathers present in the home were employed. More than a quarter of the sample families received some welfare assistance, but welfare was the primary source of income for fewer than one-sixth of the families. About half of the families in the sample had incomes under $6,000, 27 percent

*All data presented in this section are drawn from Phase III, except for information on employment of parents and sources of income, which was collected only in Phase II.

had incomes between $6,000 and $15,000 and the remainder had incomes over $15,000.*

Approximately 65 percent of all children were black, 30 percent were white, and a small fraction were of other racial origins.

Educational backgrounds of parents spanned a wide range. About 19 percent of fathers and 20 percent of mothers had not completed high school; 36 percent of fathers and 39 percent of mothers had high school diplomas only. The remaining parents had varying amounts of postsecondary education, ranging up to Ph.D. or other professional degrees. About 10 percent of fathers and six percent of mothers had bachelor's degrees or higher.

Evaluation of the Sample

NDCS centers were chosen to meet specific design requirements. They were not sampled randomly from the national population of day care centers, and therefore could not be expected to show proportional representation of all the different types of centers nationwide. In fact, the study's selection criteria guaranteed that the sample would include more than the national proportion of those types of centers of greater policy relevance (e.g., large centers, centers serving three- and four-year-olds, publicly funded centers) and less of others. However, in order to provide an adequate data base for federal policy purposes, it was not necessary that the Phase III sample show proportional representation of all of the different kinds of centers that

*Sample data on income levels were collected in Phase II, from September 1975 to September 1976. Income figures are therefore stated in 1975-76 dollars. It should be noted that a number of centers enrolled both children eligible for public subsidy and children of fee-paying parents. As far as could be determined, only eligible children were supported by Title XX funds.

exist nationally, nor even of those centers that receive federal funds. What was necessary was that the sample be sufficiently representative of centers affected by federal policy to provide an adequate basis for generalizing results to federally subsidized care across the nation. It was also important that the sample be adequate to permit detection of effects of the major policy variables--staff/child ratio, group size and staff qualifications. In sum, two questions are paramount in evaluating the NDCS sample. First, was the sample sufficiently representative of centers affected by federal policy to provide an adequate basis for generalizing results? Second, did it have the power to detect effects? Both questions can be answered affirmatively, with some minor qualifications.

The 57-center Phase II sample was compared with the Supply Study national sample on a number of dimensions--primary source of funding (government or nongovernment), number of years open; total staff size; enrollment; percent of black children enrolled; mean caregiver experience; mean caregiver education; staff/child ratio; capacity; age of oldest child; age of youngest child; and profit or nonprofit status. The comparison showed that the sample included several representatives of all types of centers, except for small, profit-making private centers. (Such centers had been deliberately excluded by the study's selection criteria because they serve few subsidized children.) In addition, the sample was reasonably representative in its distributions of the policy variables, with two exceptions. This fact was illustrated by Figures 2.1, 2.2 and 2.3, which compare distributions of the classroom composition and caregiver qualifications variables in the NDCS sample with distributions for comparable classrooms of three- and four-year-olds in the Supply Study national sample. The comparisons show highly similar profiles of means for the two samples, except for group size and experience in current center.

Mean group size in the Supply Study national sample was considerably smaller (13.8) than mean group size in the NDCS sample (17.6). Further examination of the national sample showed that small groups (12 or fewer children) tend to be found in small centers, which were deliberately excluded from the NDCS sample for reasons stated earlier. However, despite the fact that small groups were proportionally underrepresented in the NDCS sample, they were still substantially represented. Six Phase III centers had groups with average sizes of 12 or fewer children, and another 15 centers had groups ranging in size from 12 to 14. Thus the sample included enough centers to estimate costs and effects associated with groups near or below the national mean in size.

As shown in Figure 2.3B, the Supply Study national sample, compared with the NDCS sample, included proportionally more caregivers with both very large amounts of experience in their current center (more than 5 years) and very small amounts (less than 1 year). Both of these distributional facts may again be explained by the different selection criteria for the national sample and the NDCS 57-center sample: The NDCS excluded centers that had been open less than one year, while the Supply Study included such centers--obviously resulting in inclusion of proportionally more caregivers with less than one year of tenure in their current centers. Also, the NDCS excluded centers with enrollments between 15 and 25, while the Supply Study included them. In small centers, directors, who often have much more experience than other staff, frequently function as caregivers, whereas this is less common in large centers. Thus the proportion of caregivers with long experience is higher in the Supply Study sample than in the NDCS sample.

In summary, preschool classrooms in the study centers spanned the range of staff/child ratios, group sizes

and staff qualifications most relevant for policy, and they proved to be reasonably representative, with respect to those characteristics, of preschool classrooms in day care nationally.

Power to Detect Effects

The capacity to detect effects of the major policy variables depends on four characteristics of the sample: (1) adequate variation of each policy variable; (2) absence of confounding relationships among the policy variables; (3) absence of confounding relationships between the policy variables and other potential determinants of classroom processes and outcomes, such as socioeconomic characteristics of children being observed and tested; and (4) adequate size of the sample.

The first of these conditions was clearly met in both the Phase II and Phase III studies. As already indicated, NDCS centers spanned the full range of levels of the policy variables that might be embodied in federal policy decisions. Staff/child ratios ranged from current FIDCR levels to levels approximating those mandated by most states. Group sizes ranged well above and below current FIDCR levels. Staff education also varied widely--from centers with staff averaging less than a high school diploma to centers with staff averaging more than a bachelor's degree--as did staff experience, from a few months to several years. Although centers with more extreme characteristics certainly exist (e.g., centers with ratios as high as 1:3, or as low as 1:25) and while inclusion of such extreme centers would have increased the likelihood of finding effects, such extremes do not represent viable options for federal policy and were therefore excluded from study.

The second and third conditions were also largely met, although the policy variables were not completely

independent of one another, nor of background factors that also potentially affect center processes and developmental outcomes. Table 2.4 shows correlations among the major policy variables across the 57 Phase III centers.* The table indicates that the classroom composition variables are essentially uncorrelated with the caregiver qualifications variables, so that their effects can be easily separated. Within the cluster of qualifications variables, modest correlations exist--high enough to warrant caution in interpreting individual effects, but not high enough to preclude identification of the most powerful variable(s). A similar modest correlation exists between group size and staff/child ratio. Only one variable is so confounded with others as to preclude separation of its effects: number of caregivers is closely related to both ratio and to group size. (Strong links between number of caregivers, on the one hand and group size and ratio on the other are unavoidable given mathematical properties of the three variables and the distribution of these three characteristics in the day care world.)

Table 2.5 shows relationships among the policy variables and set of background variables describing the children, families and communities served by the NDCS centers. Again, many correlations are small, indicating that effects of policy variables can easily be separated from those of particular background factors. Some moderate

*In Tables 2.4 and 2.5, data for the entire Phase III sample are pooled for illustrative purposes. Most actual analyses were based on either the 49-center data base or the APS data base separately. All effects analyses were preceded by examination of correlations among independent variables in the relevant data set. Such examination was essential to check how successfully the experimental manipulation and/or balancing of independent variables had reduced confounding among these variables. In subsequent chapters, any major deviations from the overall picture shown in Tables 2.4 and 2.5 are discussed where relevant.

Table 2.4

CORRELATIONS AMONG THE MAJOR POLICY VARIABLES

(N = 57 Centers)

	Number of Caregivers	Staff/ Child Ratio	Years of Education	Child-Related Education/ Training	Previous Day Care Experience	Experience in Current Center
Classroom Composition						
Group Size	.66	−.26	−.05	.08	.04	−.14
Number of Caregivers		.56	−.00	.07	.19	.03
Staff/Child Ratio			.05	.00	.21	.19
Staff Qualifications						
Years of Education				.34	.18	−.27
Child-Related Education/ Training					.25	−.23
Previous Day Care Experience						.03

Table 2.5

CORRELATIONS AMONG POLICY AND BACKGROUND VARIABLES

(N = 57 Centers)

	Mothers' Education	Fathers' Income	Proportion White Children	Number of Adults In Home	Poverty of Surrounding Neighborhood
Classroom Composition Observed					
Group Size	.03	−.04	−.16	−.04	.00
Number of Caregivers	−.15	−.28	−.17	−.28	.25
Staff/Child Ratio	−.22	−.31	−.03	−.31	.32
Staff Qualifications					
Years of Education	.08	.14	.24	.08	−.11
Child-Related Education/Training	−.08	−.19	−.11	−.05	.30
Previous Day Care Experience	.05	−.16	.04	−.06	.09
Experience in Current Center	−.21	−.27	−.26	−.32	.38

correlations do exist, however. Perhaps most important are the associations of staff/child ratio and staff experience in current centers with various indices of socioeconomic status: high ratios and experienced staff are found in centers serving low-income families and neighborhoods, as well as children of less educated mothers, often from single-parent families. This pattern of associations is tied to federal funding. Low-income children are served in federally funded centers, which are subject to higher FIDCR ratio requirements and which pay slightly higher wages and experience lower staff turnover rates than do parent-fee centers. This pattern of relationships implies that effects of background factors, such as socioeconomic status, must be taken into account in exploring relationships between staff/child ratio or staff experience and various measures of children's behavior and development.

The final condition required for detection of the effects of the policy variables--adequate sample size--was examined statistically in planning Phase III. Computer simulation was used to estimate the likelihood that effects of varying sizes could be detected, given the projected sample size. Results of these provisional analyses, which were conducted solely for planning purposes, indicated that the sample would show detection of effects due to differences in center characteristics, as long as these effects were reasonably large relative to total variation from center to center--specifically, as long as at least 14 percent of total center-to-center variation could be explained by the policy variables. These provisional analyses were in effect confirmed by Phase III findings, which revealed many significant and systematic relationships among the policy variables, behavior and test scores.

Analytic Methods and Issues

The foregoing discussion of collinearities among independent variables leads directly to the question of statistical techniques used in disentangling the relationships between regulatable center characteristics and the experiences and development of children. Multiple regression (with covariables) was the principal statistical tool of the NDCS, augmented by a variety of analytic devices tailored to specific classes of measures. This section lays the methodological groundwork for the substantive chapters that follow, first outlining the study's general approach to regression, then discussing a series of related analytic issues that had different implications for different types of dependent variables.

NDCS Approach to Multiple Regression

The general strategy for use of regression techniques in the NDCS was an exploratory one described by a number of authors including Mosteller and Tukey.[6] This approach is oriented toward mapping complex patterns of relationships in large data sets, rather than toward rigorous testing of limited hypotheses. In this approach, a variety of regression models are explored for each dependent variable, guided by a qualitative understanding of the questions to be addressed. What is of interest is not only the individual regression coefficient or significance level resulting from a particular analysis, but also the robustness of results--the stability of estimates--across analyses. The logic of the approach is simply that a relationship that holds up across several versions of the regression model is more likely to be genuine, and less clouded by multicollinearity, than a relationship obtained once. What is sacrificed is the interpretability of significance levels; since each relationship is tested several times, no single p-value can

be regarded as meaningful in the customary sense. (In the presentation of findings in subsequent chapters, convention is honored in that t-statistics associated with various correlation and regression coefficients are reported; however, what is important, is not only the p-value associated with each t, but also the stability of t-statistics across analyses.)

Somewhat different sets of regression models were explored in each of the three domains of dependent variables.[8] This variation was motivated by several considerations. First, there were differences across the three domains in the patterns of multicollinearity among independent variables. As indicated earlier, different measures of the classroom composition variables were used in conjunction with caregiver behavior, child behavior and test scores. As a result, intercorrelations among the composition variables, and between composition and qualifications variables, occasionally deviated from the generic picture presented in Table 2.4, requiring different exploratory regression strategies. Second, preliminary analyses showed that different sets of covariables were required in the three domains. Finally, practical considerations constrained the amount of exploratory work that was possible in the three domains. For test scores, where only a few dependent measures were at issue, extensive explorations were carried out. For the domain of child behavior, where there were many measures (and where their number was in effect doubled by the need to conduct separate analyses for free play and teacher-directed activities), much less exploration was possible; after some preliminary work, essentially one model was used. The degree of exploration in the domain of caregiver behavior lay between these two poles.

Several additional comparative, exploratory analyses were carried out, again to varying degrees across the three domains of dependent variables. The principal aim of these analyses was to further establish the main effects of each of the policy variables, for main effects give the policymaker broad-brush guidance as to which regulatable center characteristics are most closely associated with the well-being of children. Some of these comparative analyses also had other policy uses, identified below. First, interaction effects were examined, to determine whether any main effects estimates were threatened. Interaction analyses also had the potential to influence the design of regulations in complicated ways. For example, certain kinds of interactions between group size and caregiver training might have suggested that group size need not be regulated for trained caregivers, but only for those with little or no training. (This is a hypothetical example; such interactions were not in fact observed.) Secondly, "biweighting*," a technique for reducing the potentially distorting effects of outlier cases, was used.[8] Third, the sample was partitioned, by center auspices and by socioeconomic status of families, in order to determine whether the overall findings held for identifiable policy-relevant subsets. (In fact, as will be seen, findings tended to be stronger for low-income children in publicly subsidized centers, the group most affected by federal policy.) Fourth, the sample

---

*Biweighted regression is an iterative procedure used to estimate the relationship between one or more independent variables and a single dependent variable. Initially, cases are assigned equal weights (corresponding to ordinary least squares) and a regression surface is estimated. Cases are then re-assigned weights that are inversely related to their distance from the fitted surface, and the regression surface is re-estimated using the new weights. The process is repeated until regression coefficients stabilize. Thus, an objective criterion is used to lessen the influence of a few possible outliers in determining the relationships between measures. Examination of the biweighted weights may also lead to the identification of outliers to be set aside in subsequent analyses.

was partitioned by site in order to determine whether effects of the policy variables held across variations in regional and local conditions. Finally, fall and spring results were compared for the child and adult observation data, as a further check on consistency. (Such comparisons were not relevant for the test data, which took the form of fall-to-spring change scores.)

Measures of State Versus Measures of Change

Fundamental decisions had to be made as to whether the study's dependent variables should be treated as state measures at a single time point or measures of change over time. In the case of test scores, the decision was relatively easy. Children enter day care with different levels of skill and knowledge, reflected in part by differences in entering scores on the PSI and PPVT. Unless the researcher controls the assignment of children to centers (a condition difficult to meet in a large-scale field study), entering skills will vary from center to center because of variation in recruitment policies and populations served.* To cite an obvious example, centers that accept all children of a given age, regardless of developmental level, are likely to have lower scores than centers that screen out children who are "not ready" for a group experience. Over time an effective center may eradicate some of the differences in relative standing reflected in entering scores, bringing children who start below the developmental level expected for their age up to the performance standards of others. However, entering differences are unlikely to be eliminated entirely. Thus, the average level of children's performance in a particular center is a dangerously misleading measure

*In the NDCS, some control over entering test scores was achieved. In the APS study, control was achieved by random assignment of children to classes. In the 49-Center study, center-average test scores from Phase II were among the variables used to match centers before assignment to "treatment" and "control" conditions.

of the impact of that center, even when measurements are made after children have been in day care for a significant period. Clearly, what is at issue is the effect of the center on the rate of change in children's scores, (or on post-test scores with entering scores taken into account--which amounts to a form of change score). However, measurement of change raises a number of difficult technical problems, which are discussed in Chapter Five.

In the case of observation measures, particularly observations of children's behavior, the proper decision was much less obvious. On one hand, it would be desirable to know how children's behavior changes over time in different day care environments. On the other hand, it is also useful to know whether regulatable center characteristics are associated with particular patterns of classroom interaction at any given time point (with confounding background characteristics of children controlled statistically). Thus a case could be made either for trying to measure fall-to-spring change in behavior patterns, or for treating the fall and spring observations as separate replications of a cross-sectional study, or both. The decision in this case was determined by practical considerations. The reliabilities of the child observation measures, though adequate for cross-sectional analysis, were too low to support analysis of change. Also, improvements in the observation procedures between fall and spring called into question the comparability of data across the two time points. (Reliabilities and observational procedures are discussed in a later section.) Consequently, observations were used as state measures. Fall and spring observations were treated as replications; primary emphasis was given to the improved spring data, and the fall data were examined for consistency and confirmation.

Attrition

Loss of participants is a problem endemic to long-term studies such as the NDCS. The problem is especially acute when dropout is selective, so that the sample changes character as well as diminishing in size over time.

The possibility of selective dropout was particularly threatening to NDCS analyses of test score gains, which depended entirely on comparability of samples within each center between fall and spring. Consider, for example, how attrition might obscure a (hypothetical) positive relationship between staff/child ratio and center-average gains on the PSI. Suppose that parents tend to remove children from centers when the children are not thriving. Suppose further that children tend to thrive in centers with high staff/child ratios. Then low-ratio centers would experience higher rates of attrition than high-ratio centers. However, children remaining in the low-ratio centers would be precisely those who, for whatever reasons, were doing well. Assuming that gain scores are one index of "thriving," this pattern of attrition would diminish the differences in gains that might otherwise distinguish high- and low-ratio centers, because children in low ratio centers who might have done poorly in spring testing would be gone when it took place. Attrition could also cloud interpretation of observation data, even though analysis of change was not planned. A change in sample composition could change the prevailing relationships between policy variables and behavioral measures, so that fall and spring data yielded different patterns of results. In such a case it would be difficult to know which data set to trust or how to compromise between the two.

However, attrition could distort NDCS findings only if the proportional loss of subjects were related <u>both</u> to one or more of the policy variables <u>and</u> to one or more

dependent variables. To pursue the above example, attrition could not mask the effects of ratio unless it occurred in low-ratio centers more than in high, and unless the children who left the sample were those who would have had low gain scores. In fact, attrition across the NDCS centers was moderate; 322 of the 1383 children (23%) tested in the fall were not tested in the spring. Moreover, rates of attrition were almost unrelated to the policy variables (see Table 2.6). Correlations were generally near zero, ranging from -.20 for child-related education/training to -.05 for years of education. (The fact that all correlations were negative is probably coincidental, but in any case it does not indicate a consistent tendency for dropout rates to be highest in centers with "worse" values of the policy variables.) For example, the negative relation with group size, -.10, indicates higher dropout rates in centers with smaller groups, i.e., in centers that were "better" in terms of the characteristic that proved to be the study's most powerful determinant of PSI gains and other benefits for children. It is of course impossible to know whether the children who dropped out of the NDCS sample between fall 1976 and spring 1977 would have had higher or lower PSI gains, or would have fared better or worse in terms of other measures. But, in the absence of strong relationships between attrition rates and the policy variables, it is unlikely that selective dropout could have distorted the study's results seriously.

Properties of Observation-Based Behavioral Measures

The NDCS relied heavily on direct observation in measuring both its dependent and independent variables. Knowledge of the metric properties of observations thus was crucial in planning the study's analyses. Most of the measurement issues surrounding observations bear on behavioral observations, such as were used to assess dependent variables in the NDCS; these are addressed in this section.

Table 2.6

CORRELATIONS BETWEEN FRACTION ATTRITED AND POLICY VARIABLES

(Center-Level Correlations; n=57)

Variables	Correlation
Group Size	-.10
Number of Staff	-.18
Staff/Child Ratio	-.12
Years of Education	-.05
Child-Related Education/Training	-.20
Previous Day Care Experience	-.18

However, some issues such as those having to do with the reliability of observation-based measures, apply both to behavioral measures and to simple head counts that were used in observing classroom composition; these are discussed in the following section.

Use of observations to study behavior in natural settings such as day care is a procedure that has strong intuitive appeal. The connection between data and phenomena is unusually direct. Natural observations avoid the artificiality that opens many laboratory studies to the charge that their findings have nothing to do with real-world behavior. Use of such observations in the NDCS exemplifies the "ecological" approach to the study of child development urged by some of the field's most prominent spokesmen, notably Urie Bronfenbrenner.[9]

Despite these advantages, observations do not give the investigator privileged access to reality. Like any measurement device, they impose their own peculiarities on the phenomena being measured. Different kinds of observation systems and different analytic approaches yield different kinds of information. Familiarity with the general properties of NDCS instruments is essential for understanding the picture of the social environment of the classroom that eventually emerged.

Use of Time-Sampled Observations

NDCS observation measures were event records, as opposed to more global ratings commonly used in studies of young children in group settings. Child observations were made on a time-sampled basis, once every 12 seconds. Caregiver behavior was recorded continuously, at the observer's own pace. (Procedural details are provided in Chapters Three and Four.)

Both time-sampled and continuously recorded observation data are sensitive to the durations as well as the frequencies of behaviors in the classroom. Such observations yield behavior profiles that are faithful to the temporal prevalence of events, and therefore are rather objective records of the experiences of children and caregivers. However, they give very little weight to events that occur infrequently or are very brief, even if these events have major psychological significance for the child or perceptual salience for the casual observer. For example, a hug or a slap may last less than a second. When such events occur, they are likely to be very important to the children involved, and memorable for an adult who happens to witness them. Yet a behavioral record of an hour-long period in which one of these events takes place will show that the event occupied a tiny fraction of one percent of the period. In contrast, more commonplace activities such as game-playing or storytelling may occupy an appreciable portion of an hour.

Because of the temporal sensitivity of time-sampled and continuous observations, numbers of recorded occurrences of individual behaviors in the NDCS varied by several orders of magnitude. Some behaviors were recorded many thousands of times in the total data set; some appeared only a few times in hundreds of thousands of records. In general, analyses concentrated on those behaviors that occurred with relatively high frequency. However in some cases where individual rare behaviors were of compelling interest, their occurrence or non-occurrence was studied using special analytic techniques. (These techniques and relevant findings are described in Chapter Five.)

Use of time-sampled observations also has the effect of producing small, but possibly important artifactual correlations among particular behaviors. Because observations

were made at more or less fixed intervals for a fixed total time span, the total number of observations was also fixed. (In the case of caregiver observations, the total number of observations varied across observers but was approximately constant across observation periods for each observer.) Consequently, if any one behavior was recorded with relatively high frequency, one or more other behaviors had to be recorded with relatively low frequency. Frequency counts for different behaviors were thus not entirely independent. Moreover, nonindependence was particularly salient for the more frequent behaviors and global construct measures created by summing frequencies of individual behaviors. As the total observation pie was cut into fewer and larger pieces, variation in the size of any one piece had increasingly noticeable affects on the amount of pie left to be split into other pieces. The mutual interdependence of observation variables was not so severe as to preclude separate analyses. However, it once again underscores the point that NDCS findings should be viewed in terms of their overall pattern and that individual effects estimates and significance levels should not be given undue weight.

Validity of Observations

If observations were used to measure traits of individual children--traits that were presumed to generalize to settings other than the day care classroom and to remain stable over time--then data drawn from the day care setting would require longitudinal cross-validation against other data sources, such as parental reports, tests, or observations in other settings. However, the NDCS used observations to assess interaction within the day care setting itself; thus issues of cross-validation did not arise.

The principal threat to the validity of NDCS observation measures was distortion of the natural behavior

of caregivers and children due to the presence of observers in the classroom. Without comparative data based on surreptitious observations of unaware caregivers and children, there is no way to know how severe such distortions were. However, observers were present in each classroom for several days, and they avoided interaction with caregivers and children. Thus there is reason to believe that the novelty of their presence may have worn off, and that gross distortions of everyday behavior due to direct contact with the persons being observed did not occur. Also, in addition to the observers, who were in study centers on a short-term basis, the NDCS employed one permanent data collector in each center for the entire two-year duration of the study. The presence of these individuals may have reduced the probability of serious alterations of normal behavior patterns during the period in which additional observers were present.

Finally, and perhaps most important, changes in behavior due to the presence of the observer would distort the study's results only if such changes were systematically related to the policy variables. Such relationships are not impossible; the tendency to alter one's behavior might be a function of one's training, or of the number of children or adults present in the classroom. However, such relationships, seem, a priori, to be less likely than global changes unrelated to the policy variables, e.g., increased attentiveness to children on the part of most or all caregivers when observers were present.

Observer Effects

Of all threats to the validity and reliability of observation instruments, the one that has received the most attention in the psychological literature is distortion of results due to differences in observer perspective.

Characteristically, considerable effort is devoted to training observers to high criteria of agreement, and often, when such standards are achieved, the researcher assumes that his or her measures are trustworthy. Although, as shown in the next section, the importance of observer effects is usually overated, and high observer agreement is no guarantee that measures are dependable, observer effects nevertheless deserve careful attention.

The first line of defense against observer effects of course lies in training. SRI recruited and trained observers carefully, and tested their performance on selected videotaped samples of behavior before and after sending them into the field. In addition, a small-scale study of inter-observer agreement under field conditions was conducted. All results indicated that satisfactory levels of agreement had been established and maintained. (Details are provided in Chapters Three and Four.)

A particularly sensitive issue having to do with observer effects arose early in Phase III, when late Phase II analyses suggested that there might exist systematic differences in perspective linked to the race of the observer. The existence of these effects could not be regarded as proven, because race of observer was partially confounded with the race of the child or caregiver under observation and with various center characteristics. Nevertheless, to guard against possible distortions due to race of observer, Phase II spring observation procedures were modified. According to the modified plan, every child and every caregiver was to be seen on successive days by two different observers, one black and one white. This modification was strongly urged by black consultants to the NDCS.[10] Despite formidable difficulties of recruitment and scheduling, SRI came close to full implementation of the plan. (See Chapters Three and Four). The procedure eliminated any

confounding between policy variables and race of observer. Moreover, it made possible a much more precise estimate of the magnitude of observer effects than would otherwise have been possible. These estimates played an important role in broader investigations of the reliability of the study's observation measures.

Reliabilities (Generalizabilities) of Observation Measures

Reliability of observation measures is an issue that can be addressed in a far more precise and satisfactory way than can their validity. Mathematical techniques for calculating reliabilities of observation data have been developed to a point of considerable sophistication. The essential ideas were set forth by Donald Medley and Herbert Mitzel as early as 1963[11] and have been most fully elaborated by Lee Cronbach and his colleagues.[12] However, these methodological advances have not yet been widely reflected in substantive work in developmental psychology.

Most researchers who use observation-based measures are content to report "inter-rater reliabilities"-- usually percentages of agreement or correlations between scores generated by pairs of observers. Less commonly, stabilities of measures across occasions of observation (usually in the form of day-to-day correlations) are also reported. Few researchers seem to be aware of the point made long ago by Medley and Mitzel, that measures of inter-observer agreement can give an extremely misleading picture of the overall trustworthiness of observation measures--even of the degree to which those measures are distored by differences in observer perspective. Moreover, not all researchers seem to recognize that, in most applications, reliabilities of

observations are threatened far more by instability over time than by observer differences.*

The approach developed by Medley, Cronbach and others integrates and generalizes the more fragmentary approaches to reliability measurement typically seen in the literature. Analysis of variance is used to estimate the components of variance in a given observation measure attributable to each important source, or "facet" in Cronbach's terminology, such as the observer, the occasion of observation, the individual child, the class or the center. Variance can then be treated as "true" or "error" depending upon the purpose of the analysis and the unit of analysis chosen. Thus a measure does not have a single reliability under this approach; rather, it has a set of reliabilities, or generalizabilities;"** in Cronbach's terms. For example, a measure of the frequency of cooperation on the part of children has one generalizability when used as a descriptor of the individual child, another when used as a descriptor of the classroom and still another when averaged to the level of the center.

Like conventional reliabilties, generalizabilties take values between zero and one, representing variance

---

*Typically the researcher wishes to use observations to characterize individual children, or classrooms, in order to relate differences among children, or differences among classes, to some other variable(s) of interest. That is, the child or classroom, not the observation, is to be the unit of analysis. Thus, typically, many observations, made at several different times, are averaged to yield a score for the child or for the class. If the child or classroom characteristic under investigation fluctuates markedly, this fluctuation reduces the reliability of the average score, even though each individual observation may be error-free.

**Cronbach's use of the term "generalizability" is not to be confused with the more conventional usage, referring to the universe to which findings based on a particular sample can be extrapolated.

ratios. The numerator of the ratio is the (estimated) amount of variance that is linked to the facet (or set of facets) of interest--e.g., child, class or center; the denominator is the total variance in that average score, which includes contributions from other sources designated as error, e.g., observer, occasion and random fluctuation. For example, a generalizability of .95 for center average staff/child ratios indicates that 95 percent of the variance in mean observed ratios is due to "true" center-to-center differences and 5 percent to nuisance variables or error, (including but not limited to class-to-class variation within centers).

It is important to note that averaging to higher levels of aggregation does not necessarily increase the generalizability of a measure. For example, if a measure is highly generalizable as an individual trait measure, but the relevant trait varies markedly within classes and does not vary systematically across classes, averaging to the class level will yield a lower generalizability than obtained at the individual level. (Child-to-child variation within classes, though quite genuine, is a source of "error" with respect to the class-average score.)

Generalizability coefficients provide two types of information that are extremely useful in approaching the analysis of observation data. First, they help in selecting the proper unit of analysis, by identifying the level of aggregation--person (child or caregiver), class or center--for which the data are most reliable. Second, they help establish the mathematical limits of the analyses to be performed--the degree of statistical power to detect relationships and the degree of bias likely to be present in estimating the strengths of relationships. When generalizabilities are modest, meaningful analyses can nevertheless be conducted if the sample provides enough degrees of freedom. However,

under such circumstances, genuine but small relationships may not reach conventional levels of statistical significance (leading to the inability to reject the null hypothesis that those relationships do not exist*).

Generalizability calculations were carried out for many of the NDCS's observation-based measures, including measures of observed group size, staff/child ratio, and qualifications of staff present in the classroom, as well as some measures of caregiver and child behavior.13 These results must be viewed as partial, rough estimates, useful primarily in planning analyses and interpreting quantitative outcomes. The findings for both independent and dependent variables may be summarizes as follows:

- In general, the occasion of observation was the dominant source of variation for all of the measures. Observer effects were much less powerful.

- Unlike public school classrooms, day care classrooms are relatively unstable partly because of absenteeism and unscheduled merging of classes and also because individual caregivers and children come and go according to idiosyncratic schedules. Thus, no single group size or staff/child ratio characterizes a classroom at all times; nevertheless, class and center averages based on multiple observations of classroom composition proved to be highly reliable descriptors of classes and centers; most reliabilities fell between .93 and .95, and none was below .8.

---

*As noted in Children at the Center, any degree of unreliability will have the effect of underestimating the bivariate relationship between two variables. In the multiple regression context generally discussed in this volume, however, it is impossible to predict the direction of change in any individual regression coefficient due to unreliability because of the effects of correlations among the independent variables entered into any specific regression equation.

- Center averages of years of education and experience of lead teachers, and center-level proportions of lead teachers with education or training in child-related fields showed only moderate generalizabilities (.3 to .6) due to fluctuations in staffing over time and variations in qualifications of lead teachers across classes. As noted earlier, these center averages were used to analyze the effects of lead teacher qualifications on test scores. (Generalizabilities of aides' qualifications were not calculated, nor were generalizabilities of class averages combining teacher and aide qualifications, which were used in analyzing effects on child behaviors.)

- Generalizabilities of construct measures describing the behavior of lead teachers were fairly high (.60 to .86) at the teacher level; that is, the variables described fairly stable behavior patterns of individuals. Since just one lead teacher was observed in each class, person-level and the class-level generalizabilities are identical for these variables. (Generalizabilities of measures of aides' behavior were not examined.)

- Generalizabilities of child behavior variables were extremely low at the child level; the variables did not appear to describe enduring traits or stable behavior of children. However, the variables showed class-level generalizabilities that were adequate for analysis, given the number of degrees of freedom involved. Generalizabilities ranged from .1 to .6, mostly clustering in the neighborhood of .3.

- Center-level generalizabilities of PSI and PPVT gain scores, calculated in a manner analogous to that used for center-level observation measures, except that "occasion" and "observer" were not relevant sources of variance, were approximately .6.

These results, together with other considerations outlined in the next section, influenced the choice of units of analysis for the NDCS. In addition, they provided a context for interpreting quantitative findings. The results suggested that certain relationships would be much easier to detect than others and that the overall explanatory power of regression models would be limited. The results implied that it would be easier to detect links between the classroom composition variables and the various dependent measures that it would be to detect relationships involving the qualifications variables. Similarly, it would be easier to detect relationships involving test scores and measures of caregiver behavior than those involving measures of child behavior. More generally, even if very strong underlying relationships between the policy variables and dependent variables were to exist, generalizability limitations would restrict the explanatory power of regression models such that even R^2's of .4 or .5 would be difficult to obtain. The larger implication was that relatively modest relationships should be taken seriously. The NDCS was a search for signals in a noisy environment; a signal loud enough to detect was likely to be stronger than it seemed against the background noise.

Units of Analysis

Data in the NDCS were hierarchically organized. Children were nested within groups or classrooms, and classrooms were nested within centers. Thus, data could be analyzed using the child as analytic unit, or data could be aggregated to classroom or center level. As already noted in the case of the caregiver, no distinction existed between the person and class levels.* However, a choice was necessary

*Behavior of lead teachers and aides was analyzed separately. Since each class had only one lead teacher, and since no more than one aide was observed in each class, the person and class levels were indistinguishable in these analyses.

between person/class and center levels. Ever since W.S. Robinson[14] showed that not only the strength but the direction of a relationship between variables can differ when examined at the individual and aggregate levels, social scientists have recognized that choice of the unit of analysis is crucial in analyzing hierarchical data. Yet there exists no general method for choosing the appropriate unit of analysis.[15]

A combination of analytic and empirical considerations led to decisions to treat measures of caregiver behavior at the person/class level, child behavior at class level, and test scores at center level. Detailed arguments justifying these decisions are presented in a paper by Judith Singer and Robert Goodrich.[16]

Singer and Goodrich note that NDCS data include three types of variables: (1) child-level variables, such as test scores, frequencies of particular behaviors, race and socioeconomic status (SES); (2) aggregate variables, such as class or center averages of test scores; and (3) global variables, such as group size, staff/child ratio and caregiver qualifications, which are defined only at class or center level and are constant for all children within a given class or center. Singer and Goodrich show that statistical estimates of the magnitudes of the effects of class or center characteristics on child behavior or test scores are identical regardless of whether analysis is conducted using the child as unit or whether an aggregate, such as class or center, is used (as long as the child-level analysis includes aggregate variables such as class-average SES, in addition to the SES of the individual child and aggregate level analyses are weighted by the number of children in each aggregate). However, significance tests based on child-level analysis yield many spurious rejections of the null hypothesis, because the tests fail to take

account of intraclass correlations arising from the fact that all children within a given class or center are exposed to the same values of aggregate and global variables describing that class or center. Singer and Goodrich conclude that the correct unit of analysis is the lowest level for which intraclass correlations do not exist. They point out that such aggregation does not entail significant loss of statistical power, despite the loss of degrees of freedom, because error variance is also reduced by taking means.

Those purely analytic considerations implied that the class or center, not the child, was the appropriate unit of analysis for child behavior measures and test scores. In the case of the child behavior measures, this conclusion was reinforced by generalizability results reported earlier, which showed that measures of child behavior were marginally reliable only when averaged to class or center level. In the case of the tests, scores were reliable at both child and center levels, but the above considerations ruled out child-level analyses. Class-level analyses were not feasible because some class enrollments were not stable over the year; children moved from class to class within centers. Thus, while center-average gain scores were meaningful, class-average scores were not.* Consequently test scores were analyzed at center level, while child behavior was analyzed at class level, to preserve as much detail as possible. In the case of measures of caregiver behavior, the person and class levels were identical, and dependent measures were reliable at that level. Hence analyses were carried out for persons/classes, again to preserve detail. All of these decisions were futher reinforced by findings on the generalizabilities of independent variables, most of

---

*The above remarks about instability of classes apply only to the 49-center study. In the more closely controlled Atlanta Public School study, classes were stable, and analyses of gain scores were carried out at class level, as discussed in Chapter Five.

which were reliable at both class and center levels and thus did not constrain choice between the two levels of aggregation.

Subsequent chapters assume familiarity with the foregoing methodological discussion. They provide substantive detail on instrumentation and procedures, and they concentrate most heavily on presentation of findings. Insofar as possible, findings have been organized to aid the reader who wishes to relate the regression results in subsequent chapters to the graphical summaries of results in Children at the Center.[17] The graphs in Children are diagrams of simple correlations between policy variables and outcome measures with one exception noted later. Diagrams were presented only for relationships which withstood testing in several regression analyses, and for which the simple correlation represents a reasonable summary. To facilitate comparison, simple correlations are included in all regression tables.

CHAPTER THREE: THE CAREGIVER IN THE CLASSROOM*

BACKGROUND

There exists a wealth of research findings which have direct or indirect application to the study of caregiver behavior in day care settings. Suggestions for potential types of caregiver behavior to be studied in the NDCS were drawn in part, from four broad areas of research: studies of how caregiver behavior is related to center characteristics; research on adult (particularly parent) behavior which promotes child development; research on teacher effectiveness with children in early grade school; and descriptions of day care environments. The available research pointed to the importance of two types of variables describing patterns of interaction between adults and children--"macro-variables" such as overall quantity of interaction with groups of various sizes, or global quality of interaction (e.g., warmth), and "micro-variables," e.g., contingent verbal response or use of rational explanations. Any or all of these macro- and micro-variables might be measured in a study of quality of day care.

The study's goals and the nature of its sample influenced the variables ultimately chosen to describe caregiver behavior in the NDCS. The NDCS operated in diverse day care settings and was chartered to examine independent variables that generally had not been studied previously. Therefore it seemed wisest to try to obtain a broad-brush picture of variations in caregiver behavior across actual day care settings, focusing on patterns of interaction assumed to be especially sensitive to classroom composition and caregiver qualifications--such as the amount

---

*This chapter is based largely on work by Barbara Dillon Goodson, reported in greater detail in Volume IV-C of the NDCS Final Report.[1] Dr. Goodson is the principal author of this chapter.

of direct interaction between caregivers and children--and on general qualitative features of caregiver-child interaction--such as active initiation of contacts with children versus more passive supervision, frequency of discipline, or amount of positive affect.

Direct observation of caregivers in day care center classrooms was the major method used to measure caregiver behavior. The instrument chosen to record behavior was the SRI Preschool Observation Instrument, or Adult-Focus Instrument (AFI). Classroom observations were conducted twice during Phase III of the study: in October 1976 and in April 1977. The observation instrument, the procedures for using it, and methods of analysis are discussed in the three sections which follow.

The Adult-Focus Instrument

The SRI Preschool Observation Instrument had previously been used by SRI International in evaluating the Follow Through and Head Start Planned Variation projects. It was modified (and hence renamed) for the NDCS to record adult behavior in day care centers. The AFI is designed to describe the day care classroom environment and to record the behavior of individual caregivers. The instrument has three sections:

- Physical Environment Inventory--a description of the equipment present in a classroom;
- Classroom Snapshot--a recording of the numbers of staff and children present at a specific point in time, and their activities and groupings; and
- Five-Minute Interaction (FMI)--a recording of the behavior of a single focus caregiver during a five-minute period.

Descriptive data from the Physical Environment Inventory were combined statistically into a single rating of physical quality for each center. (Discussion of the physical environment appears in The Classroom Environment Study in Volume IV of the NDCS Final Report.2) Classroom Snapshot data were used mainly to provide group size and staff counts for computing the classroom composition measures, while the Five-Minute Interactions (FMIs) provided the bulk of the data used in the major analyses of caregiver behavior. It is through a detailed analysis of these data in conjunction with the policy variables that the relationships between regulatable center characteristics and caregiver/child interaction were assessed.*

The FMIs were designed to provide quantitative records of caregiver behavior that had some of the form and detail of narrative descriptions. Each FMI consists of five minutes of observation, broken into 63 interaction frames. Each frame in the FMI is, in effect, a sentence about an action observed. It describes the actor (WHO), the object of the action (TO WHOM), the content of the action (WHAT), and the style (HOW). In each frame of an FMI, one code for WHO, TO WHOM, and WHAT had to be recorded. As shown in Table 3.1, there were 12 WHAT codes to choose from to indicate the action or behavior that was occurring. Because these codes are the most important in the analyses, brief definitions are provided in Table 3.2. In all observations, the focus caregiver being observed was either the

*For the spring data collection, the AFI was supplemented by a checklist completed at the conclusion of each day's observation of a classroom. The Child Development Associates (CDA) Checklist was developed and used to evaluate skills and behavior relevant to eleven functional areas of caregiver competency which have been defined in the CDA credentialing of caregivers. A detailed description of the development and content of the CDA Checklist is provided in Volume IVB of the NDCS Final Report.3

Table 3.1

PHASE III AFI CODES USED IN THE FIVE-MINUTE INTERACTIONS

WHO/TO WHOM

Teacher
Aide
Parent
Volunteer/Visitor
Child
Different Child
Toddler
Infant
Small Group (2-7)
Medium Group (8-12)
Large Group (13+)
Other

WHAT

Commands
Direct Questions
Responds
Instructs
Adult Self-Related Activity or Conversation
Center-Related Statements and Activity
Supports/Comforts
Praises/Acknowledges
Corrects
No Response
Rejects
Observes/Attends

HOW

Touch
Nonverbal
Movement

Task
Behavior
Utilitarian
Negative
Happy
Guide
Punish
Sad
Dramatic Play
Materials
Rule

Table 3.2

DEFINITIONS OF "WHAT" CODES FROM THE AFI*

COMMAND:	An order that asks for a response free of argument.
QUESTION:	Request for direct recall of material or a statement of preference.
RESPONSE:	Compliant response to a command, question, correction, or to praise.
INSTRUCT:	Demonstration of activities, explanation of rules, provision of information.
ADULT ACTIVITY:	Verbal and nonverbal activity between adults that is non-center and non-child focused.
CENTER-RELATED ACTIVITY:	Statements or activities that involve children or tasks in the center. (Examples: "Swings are fun"; adult gives each child a coloring book; adult cleans table top.)
COMFORT:	Statements or activities of affectionate attention and comfort.
PRAISE:	Approval, praise, acknowledgment, recognition, verbal or nonverbal.
CORRECT:	Attempts to change or modify a response, feeling, product or behavior.
NO RESPONSE:	A compliant response is expected but does not occur.
REJECT:	Negative, noncompliant responses, verbal or nonverbal.
OBSERVE:	Adult listens to or observes others.

*Taken from Observer's Manual, SRI Preschool Observation Instrument (Adult focus). Stanford Research Institute, Menlo Park, CA, Spring 1977.

actor (WHO) or the object (TO WHOM) of the action in each frame of the FMI. Thus the WHAT codes could include actions of the caregiver and actions directed toward the caregiver by others, especially by children. Since the observations were focused on caregivers, the caregiver was the actor in the vast majority of the NDCS data. Effects analyses were restricted to caregiver-initiated actions.

The 12 WHO/TO WHOM codes listed in Table 3.1 are basically self-explanatory. The HOW codes provide information about the action that is occurring, describing its content or affect. HOW codes were optional; none or up to six could be recorded per frame, although the average number per frame was less than 1. The relative frequencies of occurrence of the AFI codes are presented and discussed below under "Description of Caregiver Behavior."

Observers were allowed to set their own rate of coding on the FMIs. A maximum of 63 frames could be coded during each five minutes of observation, but no minimum was set. In the NDCS observations, the average number of frames completed per FMI was 54.

Phase III Samples and Procedures

Observations were conducted in all 57 NDCS study centers at two times during Phase III of the study: October 1976 and April 1977. Caregivers were observed in all classrooms that enrolled a majority of three- and four-year-old children. Two hundred ten caregivers were observed in the fall; 220 were observed in the spring.

The staff observed included both lead teachers and aides. The selection of caregivers to be observed in each target classroom followed these rules: In classrooms with

only a lead teacher, that single caregiver was observed. In classrooms with a lead teacher and aide(s), one teacher and one aide were observed. In all, twice as many lead teachers as aides were observed, approximately 140 teachers and 70 aides. Although the sample represents a near total census of the lead teachers in target classrooms, it represents only a partial sampling of aides--between one-quarter and one-third of the aides in NDCS target classrooms. Two factors account for the small proportion of aides observed: First, even if classrooms had multiple aides, only one aide was to be observed per classroom. Since the average number of aides per classroom (full- and part-time) was 2.8, only a little over one-third of the aides would have been observed even if all scheduled observations were successfully completed. Second, it was more difficult to complete observations on aides because aides were much less stable in attendance in the classrooms. Most worked part-time, and absence was much more frequent than among lead teachers. Therefore, a number of classrooms with multiple caregivers had only the lead teacher observed. For all of these reasons, results for aides are treated more tentatively than results for lead teachers in the analyses below.

In classrooms where only a lead teacher was observed, the teacher was observed for two mornings in a week. Where both a teacher and an aide in a classroom were observed, each was observed for the equivalent of a morning, usually on two days during a week. Observations of caregivers were restricted to the hours between 9 a.m. and noon, since this is the most stable period of the day in terms of child and caregiver attendance. It is also the period most linked with planned educational activities, which increased the opportunities to see caregivers interacting with children. In a morning's observation of a classroom, an average of 36 FMIs were completed.

In the fall, all observations of an individual classroom were conducted by the same observer. In the spring, however, two observers--one white and one black--were assigned to each classroom, and the focus caregiver was observed an equal amount of time by each observer. This change in procedure permitted examination of coding differences that could be attributed to an observer's race, and distributed any coding differences across caregivers and classrooms.

In both fall and spring 21 observers collected data on caregivers. Observers were selected from the local community at each site and trained by SRI International for approximately one week just before each data collection period. (A detailed description of the training may be found in SRI's Phase III Report.4) At both data collection points, observers were essentially comparable on all background characteristics except race. Most observers were female, and college graduates or soon to be college graduates; the average age was about 33 years, with observers in Detroit tending to be slightly older than the others. The primary difference between the observers hired in the fall and those hired in the spring was their race. In the fall, most observers (70%) were white, while in the spring, the number of black and white observers was almost equal in order to accommodate biracial observation teams.

Introduction to the AFI Analyses

The central AFI analyses examined the effects of the policy variables on caregiver behavior, as measured by the Five-Minute Interactions (FMIs). The first step in these analyses involved examining the frequencies and variabilities of the codes. This descriptive analysis helped set the context for analyzing the effects of the policy variables. In the descriptive analysis, all of the major

FMI codes (WHAT, TO WHOM, and HOW) were examined. In examining effects, however, only the WHAT and TO WHOM codes were used, along with two macro-codes constructed from these. The discussion of results that follows first reports the descriptive analyses and then turns to the effects analyses.

As indicated in Chapter Two, an important decision made prior to any of the analyses was the choice of the caregiver rather than the classroom as the unit of analysis. Since a teacher and an aide were observed in many classrooms, the observation data could have been combined to form classroom-level measures. Instead, however, a decision was made to examine the groups of teachers and aides separately. This approach was taken primarily because, as previously described, the aide sample was incomplete. Because some classrooms with aides had no aide data and many classrooms with multiple aides had data for only one aide, it did not seem valid to combine the data of teacher(s) and aide(s) from the same classroom.

DESCRIPTIVE RESULTS, DEPENDENT VARIABLES, AND ANALYTIC METHODS

Description of Caregiver Behavior

The FMI data shown in Table 3.3 provide a picture of the content or quality of the interactions between caregivers and children as represented by the WHAT and HOW codes. The TO WHOM codes describe the focus of the caregiver's attention.

Content of Caregiver Interactions

In terms of qualitative differences in caregiver/child interactions, the FMI WHAT codes can be organized into four broad dimensions: 1) SOCIAL INTERACTION, involving

Table 3.3

CAREGIVERS' ACTIONS TOWARD DIFFERENT RECIPIENTS:
MEAN PROPORTIONS (STANDARD DEVIATIONS) OF WHAT AND TO WHOM CODES, SPRING AFI
(n=220)

	TO WHOM Code					
WHAT Code	1 child	(2-7) Small Group	(8-12) Medium Group	(13+) Large Group	All Children	Staff and Environment
	$\bar{x}$ (s.d.)	$\bar{x}$ (s.d.)	$\bar{x}$ (s.d.)	$\bar{x}$ (s.d.)	$\bar{x}$ (s.d.)	$\bar{x}$
Commands	.057 (.028)	.009 (.008)	.008 (.014)	.009 (.014)	.083 (.038)	
Corrects	.052 (.028)	.006 (.005)	.003 (.008)	.003 (.006)	.064 (.035)	
Instructs	.022 (.024)	.016 (.027)	.022 (.007)	.021 (.039)	.081 (.060)	
Questions	.044 (.028)	.005 (.008)	.005 (.011)	.004 (.009)	.058 (.034)	.003
Response	.016 (.013)	.000	.000	.000	.016 (.013)	.003
Comforts	.012 (.014)	.000	.000	.000	.012 (.014)	
Praises	.038 (.026)	.002 (.003)	.002 (.005)	.002 (.005)	.043 (.029)	
Center-related	.058 (.042)	.010 (.014)	.007 (.013)	.008 (.020)	.084 (.054)	.289
Adult-related	.000	.000	.000	.000	.000	.050
Observes	.024 (.027)	.048 (.055)	.046 (.062)	.085 (.112)	.203 (.114)	.007
TOTAL	.323 (.125)	.096 (.077)	.093 (.110)	.132 (.141)	.644 (.145)	.352

positive caregiver/child interactions (usually involving caregiver verbalization), both directive and nondirective; 2) MANAGEMENT, involving caregiver/child interactions focused on amending children's behavior; 3) OBSERVATION/SUPERVISION, when the caregiver stands back and watches children; and, 4) CENTER- OR ADULT-RELATED BEHAVIOR, mostly relating to caregiver actions in which children are not focal. Although in theory the code for center-related activity could involve interaction with children or materials (see Table 3.2), in the NDCS observations most center-related activity was not directed at children. Thus, in this study, center-related activity largely represents non-child activity.

The first two dimensions above, social interaction and management, represent active engagement with children; together they accounted for an average of 37 percent of a caregiver's time. The latter two dimensions represent non-interactive behavior and occupied, on the average, over half of a caregiver's time (Table 3.3). In particular, an average of 20 percent of a caregiver's time was spent observing/attending children, and 34 percent was spent in either adult-related activity or center-related activity not involving children.

All of the codes representing verbal interaction with children--INSTRUCTS, RESPONDS, PRAISES, COMFORTS, QUESTIONS, COMMANDS and CORRECTS--were positively correlated with each other, and negatively correlated with the codes representing passive caregiver behavior with children --OBSERVES, CENTER ACTIVITY, and ADULT ACTIVITY (Table 3.4). Among the social interaction codes, instructing occupied eight percent of the caregiver's time. Thirteen percent of caregiver time was spent "warmly" interacting with children --praising, comforting, asking questions of and responding

to children, a set of codes that were highly correlated. (Note that the codes COMFORTS and RESPONDS were particularly infrequent.) An additional 15 percent of the caregiver observations were coded as COMMANDS or CORRECTS, representing efforts to alter behavior, manage or control children. These two codes also were strongly correlated (Table 3.4).

The 20 percent of a caregiver's time spent observing/attending children was approximately twice as much as any other single caregiver activity with children. As recorded in the NDCS, the code OBSERVES appears to have reflected passive supervision of children. Observing is not inherently passive, but the pattern of correlations among the WHAT codes suggest that, within the range of frequencies observed in the NDCS, more observing meant less of almost all other activities with children. OBSERVES was negatively correlated with all of the other codes except ADULT ACTIVITY. Although intelligent observation of children is a hard-won skill of the trained caregiver, the instrument did not distinguish different types of observing by caregivers.

An average of a third of a caregiver's time was spent in activities that did not involve interaction with or observation of children. Most of this time was spent in center-related activity, such as preparing or passing out materials. Only 5 percent of a caregiver's time, on the average, was spent in dealings with other adults.

The pattern of caregiver behavior that emerged was strikingly similar in quality and quantity for the fall and spring observations. At both time points, teachers and aides behaved somewhat differently in the classroom.

Table 3.4

INTERCORRELATIONS OF WHAT AND TO WHOM CODES, SPRING AF1

(n=220)

	COMMAND	CORRECT	QUESTION	INSTRUCT	RESPOND	PRAISE	COMFORT	OBSERVE	ADULT ACTIVITY	CENTER ACTIVITY	TO ONE CHILD	TO SMALL GROUP	TO MEDIUM GROUP	TO LARGE GROUP	TO STAFF	TO ENVIRONMENT
COMMAND		.41	.30	.20		.27		-.19	-.29	-.37	.46		.17		-.17	-.45
CORRECT			.20	.14	.16	.16		-.22	-.23	-.30	.43			-.13		-.34
QUESTION				.29	.22	.50		-.26	-.27	-.37	.61			-.19	-.17	-.43
INSTRUCT					.17	.27		-.37	-.16	-.33	.19		.23		.19	-.27
RESPOND						.43	.19	-.28	-.17		.47	.18		-.22		-.25
PRAISE								-.28	-.32	-.31	.56			-.17		-.42
COMFORT								-.23			.34			-.20		
OBSERVE										-.42	.29			.48		-.27
ADULT ACTIVITY											-.39	-.27			.26	.26
CENTER ACTIVITY											-.32		-.27	-.27	.17	.78
TO ONE CHILD												.29		-.38		-.60
TO SMALL GROUP													-.25	-.37		-.20
TO MEDIUM GROUP														-.38		-.21
TO LARGE GROUP															-.20	
TO STAFF																
TO ENVIRONMENT																

Note: Correlations reported are p<.05. Correlations above .18 are significant at p<.01.

Compared to teachers, aides did less commanding and instructing, and more observing (Table 3.5). This pattern is understandable, since aides in the NDCS classrooms typically acted as assistants with less responsibility than the lead teacher.

Who Caregivers Interacted With

We can expand the broad picture of caregiver behavior gained from the WHAT codes by studying the recipients of the caregivers' attention. Approximately one-third of caregivers' behavior (including both observation and more active focus of behavior) was directed toward individual children, one-third toward groups of children and the remaining one-third either toward other staff or toward the physical environment. Of the behavior directed toward children, about half was directed toward individuals, while the remaining half was about equally divided among small, medium and large groups of children. Teachers and aides showed very similar distributions of their attention (Table 3.5).

What caregivers did and whom they worked with were strongly related. The joint distribution of WHAT and TO WHOM codes suggests that different kinds of activities occurred with different numbers of children (Table 3.3). (This is also borne out in the correlations of the WHAT and TO WHOM codes shown in Table 3.4). When caregivers were instructing, they were as likely to be involved with more than one child as with individual children. Other activities occurred nearly exclusively with individual children: QUESTIONS, RESPONDS, COMFORTS, and PRAISES. These were "warmer" and more interactive codes. COMMANDS, CORRECTS, and CENTER-RELATED ACTIVITY occurred mostly with individual children but also with groups. The code OBSERVES was in a

Table 3.5

MEAN FREQUENCIES OF WHAT AND TO WHOM CODES AS A FUNCTION OF CAREGIVER JOB, SPRING AFI (n=173*)

	Teachers (n=115)	Aides (n=58)	Significance Level of Difference
WHAT Codes			
COMMANDS	.086	.070	.01
QUESTIONS	.061	.059	
RESPONDS	.022	.019	
INSTRUCTS	.090	.068	.01
ADULT-RELATED ACTIVITY	.060	.033	
CENTER-RELATED ACTIVITY	.380	.380	
COMFORTS	.015	.013	
PRAISES	.047	.047	
CORRECTS	.066	.064	
OBSERVES	.172	.240	.00
TO WHOM Codes			
TO 1 CHILD	.341	.341	
TO SMALL GROUP	.082	.117	.05
TO MEDIUM GROUP	.091	.077	
TO LARGE GROUP	.109	.118	
TO CHILDREN	.634	.653	
TO STAFF	.057	.064	
TO ENVIRONMENT	.278	.266	

*Caregivers from the Atlanta Public School centers were not included because of manipulations of job functions made as part of NDCS.

class by itself; it became more frequent as the size of groups increased and was usually recorded between a caregiver and a large group of children. Caregivers observed/supervised larger groups of children during their free play periods; observation of smaller groups occurred in both free play and during task-oriented activities where the caregiver had structured the activity and then let the children work on their own.

How Caregivers Interacted

The remaining set of FMI codes--the HOW codes--described the manner in which caregivers interacted with children. All of the HOW codes were recorded infrequently, however, since they were optional; and therefore the codes were not analytically useful. Only about half of the codes had frequencies above .01 (Table 3.6). Further, the HOW codes with the highest frequencies, such as MOVEMENT, were of least substantive interest, while those most closely tied to theoretical concepts were rare events.

Caregiver affect was of some interest. Overt affect--NEGATIVE or POSITIVE--was coded relatively rarely; however, POSITIVE affect was recorded more than three times as often as NEGATIVE. When the categories of POSITIVE affect and TOUCH are combined, it is clear that some positive interaction occurred in approximately eight percent of a caregiver's observations. The indicators of positive and negative affect usually accompanied direct caregiver-child interchanges. POSITIVE affect was coded most often in the context of praising. Caregivers touched children most often while comforting them. Not surprisingly, NEGATIVE affect was exhibited most often when caregivers corrected children. In fact, about 25 percent of the time that CORRECT was coded,

Table 3.6

MEAN PROPORTIONS OF HOW CODES, SPRING AFI

(n=220)

	x	s.d.
TOUCH	.036	.03
NONVERBAL	.275	.12
MOVEMENT	.181	.09
TASK	.100	.08
RESPONSE TO CHILD BEHAVIOR	.051	.03
UTILITY	.131	.09
NEGATIVE, PUNISH	.008	.01
POSITIVE, HAPPY	040	.07
GUIDE	.008	.01
SAD	.000	
DRAMATIC PLAY	.003	.01
MATERIALS	.028	.04
RULE	.004	.01
NO RESPONSE TO CHILD BEHAVIOR	.000	

it involved NEGATIVE affect by the caregiver; moreover, the majority of the caregivers' corrections were responses to children's behavior (or misbehavior).

Selection and Construction of Dependent Measures

The dependent measures in the effects analyses included all of the WHAT codes and the TO WHOM codes which occurred with frequency above .01. The HOW codes were rejected because of their low frequencies of occurrence and badly skewed distributions. In addition to the individual WHAT codes, two macro-codes were constructed and used as dependent measures.

Several strategies were used in an attempt to find patterns of caregiver behavior among the individual FMI codes that could be represented in constructs or macro-codes. The first technique used was a principal components factor analysis of the data, which revealed little underlying structure (i.e., no stable factors). The first factor derived in the factor analysis accounted for less than 15 percent of the variance; no other factor accounted for more than 10 percent. The first factor presented almost exactly the same picture as the simple correlations: ADULT-RELATED ACTIVITY, CENTER-RELATED ACTIVITY, and OBSERVES had negative weights while the remaining codes had high positive weights (with the exception of COMFORTS, which had a loading of essentially zero). However, since this principal component accounted for relatively little variance, no single summary construct was formed from the FMI codes.

The lack of guidance from the factor analysis led back to the raw frequencies of the codes and their correlations, which were interpreted with the help of an empirical understanding of behavior in day care settings. Based on the correlations (discussed in the descriptive

analyses) and the conceptual relations among codes, two constructs were formed. One, labeled MANAGE, was a combination of the two highly correlated codes COMMANDS and CORRECTS. This construct represented caregiver efforts to change or control children's behavior. The second construct, SOCIAL INTERACTION, was formed by combining QUESTIONS, RESPONDS, INSTRUCTS, PRAISES and COMFORTS. The SOCIAL INTERACTION construct represents all verbal social interactions between caregivers and children, excluding managing children.

The discussion of results that follows focuses on a subset of the dependent measures: all of the TO WHOM codes, the individual WHAT codes, CENTER ACTIVITY and ADULT ACTIVITY, and the two constructs SOCIAL INTERACTION and MANAGE. These exclude the individual WHAT codes that comprise the constructs. Results for all of the codes are provided in a fuller report on the AFI in Volume IV of the NDCS Final Report.

Reliability of the Dependent Measures

The reliability of the AFI measures was assessed in three ways: generalizability computations, observer agreement (with criterion tapes and in tests of interobserver agreement), and examination of the stability of the measures across timepoints. The reliability analyses indicated that the measures were sufficiently reliable to support the effects analyses, that is, we could expect a significant part of the variance in the measures to be systematic and potentially explainable by the policy measures. On the other hand, the measures were not so reliable as to predict that more than moderate amounts of variance would be accounted for. The reliability analyses also indicated that the broader dependent measures, especially the macro-codes, would be best predicted.

The generalizability computations are discussed in Chapter Two. The results of the observer tests are reported here. Observer effects were examined by SRI International through observer agreement with criterion videotapes and a field-tested interobserver (paired) agreement.5 On the criterion videotapes, agreement on all AFI codes was above 70 percent. In the field test of interobserver agreement, observer pairs of one black and one white member observed caregivers in the Five-Minute Interactions, spaced a week apart. Rates of agreement were approximately 90 percent for WHO and TO WHOM codes. Agreement varied from 62 to 89 percent for the frequent WHAT codes that were used in effects analyses, with most of these codes in the 70-85 percent range. HOW codes in many cases produced high percentage agreement, based on very few occurrences. Black and white observers differed in their use of certain codes; however, many of these differences were attributable to one or two observer pairs or to low overall frequencies of the codes in question. On the whole, SRI's data suggest that interobserver agreement, while far from perfect, is good enough to guarantee that recorded frequencies of AFI codes are determined mainly by factors outside the eye of the beholder.

Day-to-day stabilities of code frequencies were examined for 203 caregivers who were observed on two consecutive days in spring 1977. Stability coefficients, shown in the first column of Table 3.7, are correlations between frequencies of the same code measured on successive days for the same caregivers. Modest correlations were obtained --generally around .2. These indicate some tendency for profiles of caregiver behavior to remain the same, but they also show that behavior fluctuates in response to the situation, with many caregivers showing a lot of a given kind of behavior on one day, followed by relatively little on the next day. (Low values of coefficients in Table 3.7

Table 3.7

STABILITIES OF ADULT-FOCUS DEPENDENT MEASURES

Adult-Focus Codes/Constructs	Day-to-Day Stability (Spring 1977; n=203 caregivers)	Fall-to-spring Stability (Phase III; n=145 caregivers)
TO WHOM Codes		
TO ONE CHILD	.28	.26
TO SMALL GROUP	.24	.36
TO MEDIUM GROUP	.31	.26
TO LARGE GROUP	.40	.40
TO OTHER STAFF	.40	.14
TO ENVIRONMENT	.06	.24
WHAT Codes		
COMMANDS	.13	.14
CORRECTS	.06	.07
QUESTIONS	.16	.27
RESPONDS	.14	.49
PRAISES	.20	.47
COMFORTS	.19	.07
INSTRUCTS	.07	.36
OBSERVES	.32	.38
ADULT-RELATED ACTIVITY	.25	.36
CENTER-RELATED ACTIVITY	.06	.26
Constructs		
MANAGEMENT	.14	.27
SOCIAL INTERACTION	.22	.37

are to a degree artificial, since changing observers from one day to the next, required by the spring data collection plan, contributed to the apparent instability of codes. However, in light of the relatively high rates of interobserver agreement obtained in SRI's field test and the results of the generalizability calculations reported in Chapter Two, the relatively weak correlations shown in the table must be attributed primarily to volatility of caregiver behavior, rather than to observer differences.)

Fall-to-spring stabilities, shown in the second column of Table 3.7, were with a few exceptions approximately as high as day-to-day stabilities and in a few cases were substantially higher. (Correlations in the table are based on a sample of 145 caregivers observed in both fall and spring. Scores for each caregiver were averaged over two days of observation at each time point.) The fact that long-term stabilities do not deteriorate suggests that there is some long-term continuity in caregiver behavior as measured by the AFI. In some cases this continuity is partially obscured by short-term fluctuation.

The overall pattern of stability coefficients is a mix. Where there are low stabilities at both points, this suggests that the immediate situation controls behavior, rather than any characteristic of the caregiver. Low stabilities in the short term, together with higher long-term stabilities, suggest that there are general and long-lasting caregiver styles, but that these may be hard to detect over a short span of observation because day-to-day changes in the situation inhibit expression of the caregiver's usual dispositions. Altogether, the results of the stability analyses suggest that a good part of the variance in the measures of caregiver behavior may <u>not</u> be systematically related to fixed characteristics of the caregiver or the classroom. (And, since the policy measures are fixed

characteristics of this type, the results indicate that the strength of potential relationships between policy variables and caregiver behavior may be limited.)

Regression Analyses

The goal of the main effects analyses was to define the relationships between each independent policy measure and caregiver behavior, and to assess how well the set of policy variables predicted caregiver behavior. Data were analyzed by multiple regression, using different combinations of the policy variables, selected so as to minimize the confounding among the set and maximize the chance of statistically separating the effects of the policy variables.

Regression Models

Ten independent measures were the focus of the AFI effects analyses. Eight were policy variables: observed group size, number of staff and ratio of staff to children, caregiver years of education, child-related education/training (also called specialization), previous day care experience, experience in current center, and age; two were covariables--caregiver race, and socioeconomic status (SES) of the children.* The ten were not entered as a single group in any of the regression equations for two reasons. First, the set was too large relative to the sample sizes of the data sets. Second, there were problems of multicollinearity among the independent measures.

*The variable for SES of the classroom was a construct representing five measures: parent education, family size, family income, number of parents, and race of child. The five variables were factor analyzed and a principal component factor score was assigned to each class.

Two multicollinearities were particularly salient in the AFI data. First, the correlation of group size and ratio was relatively high ($r = -.45$ for teachers and $-.65$ for aides). Second, ratio was correlated with experience in current center ($r = .41$). In addition to these confoundings, the AFI data shared with the other NDCS data samples the confounding between years of education and specialization, ($r = .38$). (See Table 2.4, earlier in this volume, which presents a generic correlation table for NDCS samples.) Finally, race of caregiver was correlated with the average SES level of the classroom.

As a result of these confoundings, several hierarchical regression models were employed with the AFI data, using different independent policy measures in each model. The two covariables were entered first in every model. Then, the set of policy measures in each model were entered stepwise. Only the final step of each regression is reported, since there was no theoretical basis for predicting or interpreting the order of entry of the policy measures, and since coefficients are not affected by order of entry.

Three regression models are discussed in detail in this chapter (listed as principal models in Table 3.8). Model I entered two policy variables which were not confounded: group size and child-related education/training (hereafter called "specialization" for brevity). In Model II, ratio was entered along with specialization. Regression Model III entered the variables for experience in current center, group size, and specialization. Two covariables--caregiver race and classroom SES--were entered in every model. Three secondary regression models are also listed in Table 3.8. One model entered both ratio and group size together with specialization, and the other two models

Table 3.8

OUTLINE OF AFI EXPLORATORY REGRESSIONS

Model	Covariables	Policy Variables	Purpose
		PRINCIPAL MODELS	
I	Class SES Caregiver Race	Group Size Specialization	• Estimate individual and combined effects of GROUP SIZE and SPECIALIZATION
II	Class SES Caregiver Race	Ratio Specialization	• Estimate individual and combined effects of RATIO with SPECIALIZATION
III	Class SES Caregiver Race	Group Size Specialization Experience in Current Center	• Estimate effects of caregiver EXPERIENCE
		SECONDARY MODELS	
	Class SES Caregiver Race	Group Size Ratio Specialization	
	Class SES Caregiver Race	Group Size Years of Education	
	Class SES Caregiver Race	Group Size Specialization Previous Day Care Experience	

tested years of education and previous day care experience, respectively. These last two models are not discussed further because the policy measures of interest had no effects. The first of the secondary models highlights the interpretive difficulties created by the collinearity of ratio and group size; results from this model are discussed in conjunction with results of primary models I and II.

Several considerations motivated this choice of regression models: (1) There was no reason to try to separate the effects of caregiver race and class SES. Therefore the two measures were entered simultaneously into all regressions, and only their combined effect was examined. We assumed that there was a "package" of caregiver and child background factors that was likely to be related to caregiver behavior and that should be taken into acount.

(2) Unlike the case of the covariables, assessing and comparing the individual effects of group size and ratio on caregiver behavior was of central interest; their confounding, however, made this impossible. Entering the two together in the regression models was problematic for interpretation because of their multicollinearity. Entering them separately would not disentangle their effects; any group size effect could also be interpreted as a ratio effect, and vice versa. We chose the strategy of trying both approaches--entering group size and ratio together and entering them separately. (In the following discussion, the focus is on the results for the separate models, for two reasons. First, group size and ratio were shown to be related to somewhat different caregiver behaviors in a systematic way that suggested the two composition measures were confounded but not synonymous. Second, the regression model with both variables entered produced some artifacts, either spurious effects for one or the other or no effects for one when its sample correlation with the dependent measure was high.)

(3) Among the qualifications variables, specialization was of primary interest because it showed many significant simple correlations with adult behavior variables and because it was a significnt predictor of test scores and child behaivor. (See Chapters Four and Five.)

(4) Education was studied to a limited extent because it was not a strong predictor in other domains and its simple correlations with lead teacher behavior were weaker than those of specialization. (Among aides, education was about as good a predictor as specialization, but neither was very powerful.) Also, education tends to be an SES measure. Education was correlated with race of caregiver, while specialization was not. (The decision to de-emphasize this variable was supported when regression results indicated no significant effects for education.)

(5) The caregiver experience variables were tested in models along with group size and specialization. They were not tested in models with ratio, because experience and ratio were confounded in the AFI sample.

Tests for Robustness

In addition to the conventional least-squares regression analyses, two kinds of checks were done on the AFI data to identify extreme or atypical cases which would play havoc with the distributions required for the kinds of statistics used in the analyses. First, scatterplots of the dependent and independent measures were scanned for bivariate outliers. Second, biweighted regressions were run, to assess the robustness of the regression equation if outliers are removed. In biweighted regressions, weights are assigned to cases on the basis of their deviations from the regression surface. Outliers are given less weight and thus affect the regression equation less strongly (see Chapter 2 of this volume).

The scatterplots clearly indicated a handful of about seven outlier cases. These cases either had extreme values on the dependent measures (e.g., 65 percent on OBSERVES when the next highest value was below 50 percent), or extremely small values for group size. Once these cases were eliminated from the data set, the correlations between the policy and dependent measures were calculated. Those that were significant before the exclusion became stronger, and some of the apparently contradictory and/or unexplainable correlations disappeared. In general, however, the main effects of the policy variables were not dependent on the few atypical cases with extreme values.

Further, the estimates obtained with the biweighted regression analyses typically were not substantially different from the estimates from the unweighted regressions. Biweighted regressions were done on the spring data for the 49-center lead teachers. In general, the biweighted estimates for the regressors were similar to the unweighted estimates. Estimates for group size were virtually unchanged by the weighting, and a few of the estimates for ratio were reduced.

Samples in the Regression Analyses

Observation data were collected on teachers and aides in the 49 centers and in the APS centers, at fall and spring. The regression models were investigated separately for the following samples:

- 49-center teachers--spring;
- 49-center aides--spring;
- APS teachers--spring;
- APS aides--spring;
- 49-center teachers--fall; and
- 49-center aides-fall.

This report focuses on the spring data for lead teachers in the 49-center sample. Spring data are emphasized because the data collection techniques in the spring controlled better for observer effects and because of the instability of caregivers' classroom assignments in the fall. Data on teachers are emphasized because of the representativeness of the teacher sample and its greater policy interest.

Following discussion of effects of each of the policy variables in the spring data for 49-center lead teachers, the consistency of the findings across the other samples is discussed briefly. In addition, the consistency of findings is discussed in various stratified subsamples of the 49-center lead teacher data. Regression Model I (group size and specialization) was examined for the spring lead teacher data stratified in four different ways: by site, by center auspice (private vs. public sponsorship), by center funding source (some federally funded children enrolled or none enrolled), and by income level of the center population (low or medium income).

While these comparisons and consistency checks contribute much valuable information, there are several points to be kept in mind in interpreting their results: First, the sizes of most of the stratified samples within the 49-center study are small enough to reduce statistical power to detect effects. Therefore, when a significant effect in the 49-center teacher sample is matched by non-significant findings in sub-samples, as is often true, it cannot be determined whether the nonsignificance is because of a null finding or a lack of power to detect the effect. Consistency of direction is as important to consider as significance levels.

Second, comparison between the 49-center and APS results is of interest primarily for the light it sheds on the classroom composition variables. The APS centers

provided a strong test of the effects of these variables, because of the homogeneity of the child population and caregivers in the centers (all were black and similar in SES) and because there was random assignment of children to classes within centers during Phase III. However, for the AFI data, APS and 49-center results could not be compared for the caregiver qualification variables. The APS centers were used for an experiment to test the effects of level of caregiver education, which involved promoting some aides to teachers and demoting some teachers to aides. Consequently, the terms "teacher" and "aide" had different meanings in APS centers. Further, the resulting groups of aides and lead teachers had different profiles of qualifications from the 49-center teachers and aides. Also, because of a unique training program in Atlanta, most of the caregivers had child-related education. These differences make it difficult to interpret comparisons between the APS and the 49-center results. Different effects might arise because the policy variables operate differently in APS centers or because the promoted aides in the APS sample behaved more like aides than like teachers.

Finally there are multiple possible explanations for differences that might be found between the fall and spring results. Fall-spring differences might reflect actual differences in relations between policy variables and caregiver behavior at different times of year; for example, group size might operate differently in October, when centers are getting organized and integrating new children, from April, when acquaintenceships and social patterns are established. However, methodological factors could also account for fall-spring differences. The different data collection procedures might be responsible for differences in outcomes. (As noted earlier, in the spring each caregiver was observed on two mornings, by two observers, one black and one white. In the fall, each caregiver was observed on

one or two mornings, by one observer, usually of the same race. Thus, the fall results are more likely to be confounded with observer differences, particularly observer race.) Fall-spring differences also might result from fallible measures, both independent and dependent. Finally, there were some notable differences in the fall in the correlations among the independent measures, particularly for aides. Among aides, there was more confounding among the independent measures at fall than at spring, particularly between the experience variables and the classroom parameters.

EFFECTS OF THE POLICY VARIABLES

Lead Teacher Behavior in the 49-Center Study

Most reported findings for lead teachers are based on a sample of 87 teachers. (Of the 115 teachers observed in the 49-center study, these 87 had no missing data on the background variables used as independent measures in the regression.) In the presentation of the regression analyses for this sample, the results are organized around the major independent measures. Findings for the group composition measures (group size and ratio) are discussed for each dependent measure first, followed by discussion of findings for caregiver qualifications, and finally, the covariables. The discussion is accompanied by tables of regression results--one table for each of the dependent measures. For each dependent measure, the table presents the findings from all of the regression models.

Group Composition Measures

SOCIAL INTERACTION (Table 3.9). There was a tendency for positive social interactions between lead caregivers and children to take place more frequently in small groups than large. The relationship is significant

Table 3.9

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

SOCIAL INTERACTION

Policy Variables		Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	-.003	1.42	.17	-.14	.08
	Child-related education/ training	.052	2.56	.01	.27	(.17)
II	Observed staff/child ratio	-.110	.62	.54	-.02	.07
	Child-related education/ training	.052	2.54	.01	.27	(.16)
III	Observed group size	-.003	1.27	.21	-.14	.09
	Child-related education/ training	.055	2.58	.01	.27	(.19)
	Experience in current day care center	.000	.14	.88	.06	

($p = .05$) in the simple regression of SOCIAL INTERACTION on group size for the sample of 115 lead caregivers observed in the 49-center study.* In the model reported in Table 3.9, for which $n = 87$, the relationship is no longer significant, although the direction of the relationship persists.

MANAGEMENT (Table 3.10). Both group size and staff/child ratio were related to the amount of management by caregivers. Larger group sizes tended to accompany more managing of children by the caregiver, while higher staff/child ratios were associated with less managing by lead teachers. The relationship between staff/child ratio and MANAGEMENT is particularly strong.

OBSERVES (Table 3.11). The amount of time that a teacher spent observing but not actively involved with children was strongly related to both the number of children and to the staff/child ratio in the classroom. Lead teachers in larger classrooms tended to do more observing; conversely, teachers in higher ratio classrooms tended to do less observing.

ATTENTION TO ONE CHILD; SMALL, MEDIUM, AND LARGE GROUPS (Tables 3.12-3.15). Neither staff/child ratio nor group size was related to the amount of time that lead teachers spent interacting with individual children. However, both variables were related to patterns of group interaction.

Group size was a strong predictor of how caregivers distributed their attention in the classroom: as group size increased, teachers spent less time with small

*This result was reported in Volume I of the NDCS final report, Children at the Center. (See Preface references for full citation.)

Table 3.10

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

MANAGEMENT

Policy Variables		Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	.002	1.65	.10	.17	.03
	Child-related education/ training	-.003	.22	.83	-.02	(.06)
II	Observed staff/child ratio	-.347	3.02	.003	-.30	.09
	Child-related education/ training	.002	.14	.90	-.02	(.12)
III	Observed group size	.002	1.17	.25	.17	.05
	Child-related education/ training	-.003	.22	.83	-.02	(.08)
	Experience in current day care center	-.003	1.33	.19	-.20	

Table 3.11

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

OBSERVES

Policy Variables		Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	.006	3.18	.002	.32	.10
	Child-related education/ training	-.022	1.25	.22	-.11	(.26)
II	Observed staff/child ratio	-.386	2.52	.01	-.28	.07
	Child-related education/ training	-.014	.78	.44	-.11	(.23)
III	Observed group size	.006	3.03	.003	.32	.11
	Child-related education/ training	-.021	1.15	.25	-.11	(.27)
	Experience in current day care center	.000	.10	.94	-.04	

Table 3.12

RESULTS OF REGRESSIONS OF CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

ATTENTION TO ONE CHILD

Policy Variables		Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	.001	.42	.68	.04	.01
	Child-related education/ training	.026	.89	.37	.09	(.10)
II	Observed staff/child ratio	-.264	1.10	.28	-.08	.02
	Child-related education/ training	.026	.94	.35	.09	(.11)
III	Observed group size	.001	.40	.67	.04	.01
	Child-related education/ training	.026	.89	.38	.09	(.11)
	Experience in current day care center	.001	.24	.81	-.02	

Table 3.13

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

ATTENTION TO SMALL (2-7) GROUPS

Policy Variables		Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	-.004	2.90	.01	-.29	.10
	Child-related education/ training	.009	.63	.53	.02	(.11)
II	Observed staff/child ratio	.345	2.75	.007	.29	.08
	Child-related education/ training	.001	.10	.94	.02	(.09)
III	Observed group size	-.004	2.50	.02	-.29	.13
	Child-related education/ training	.004	.24	.81	.02	(.14)
	Experience in current day care center	.004	1.61	.11	.23	

Table 3.14

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

ATTENTION TO MEDIUM GROUPS

	Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	-.115	2.38	.02	-.26	.09
	Child-related education/ training	-.029	1.53	.13	-.13	(.14)
II	Observed staff/child ratio	-.444	2.79	.007	-.31	.10
	Child-related education/ training	-.026	1.41	.16	-.13	(.16)
III	Observed group size	-.006	3.13	.002	-.26	.18
	Child-related education/ training	-.018	.98	.33	-.13	(.23)
	Experience in current day care center	-.009	3.22	.002	-.23	

Table 3.15

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

ATTENTION TO LARGE (13+) GROUPS

Policy Variables		Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	.013	5.74	.000	.54	.30
	Child-related education/ training	.031	1.40	.17	.16	(.31)
II	Observed staff/child ratio	-.608	2.81	.006	-.28	.12
	Child-related education/ training	.050	2.00	.05	.16	(.14)
III	Observed group size	.014	5.63	.000	.54	.31
	Child-related education/ training	.030	1.30	.20	.16	(.32)
	Experience in current day care center	.002	.69	.50	-.05	

groups (2-7) and more time with larger groups (13 or more). These relationships are not trivial or tautological, as they may appear at first glance. The independent variable "group size", referring to the total number of children supervised by a caregiver or team of caregivers, is not the same as the dependent variable describing the number of children toward whom the caregiver directs her attention at a particular moment. For example, it is possible for teams of caregivers in charge of large groups to form smaller activity subgroups, so that measures of caregivers' attention would show little or no relationship to total group size. The data, however, suggest that this sort of division is not the norm, although it does occur. Lead teachers spend a significant portion of their time interacting with most or all of the children for whom they are responsible; the larger the total group, the more their attention is spread. (This relationship holds even when classes of 12 or smaller are excluded from the sample, eliminating all cases for which total group size imposes a tautological constraint against interaction with large groups.)

Staff/child ratio also was strongly related to the number of children with whom the teacher interacted. Lead teachers in higher ratio classrooms spent more time with small groups of children and less time with medium and large groups. (They also spent more time with other staff, as discussed below.)

Non-Child Activities: CENTER-RELATED ACTIVITY, ATTENTION TO STAFF, and ADULT-RELATED ACTIVITY (Tables 3.16-3.18). Only staff/child ratio, not group size, was related to the amount of time teachers spent not involved with children. A higher ratio of staff to children (which usually implied more staff) meant that teachers spent more time in tasks which did not directly involve children--e.g.,

Table 3.16

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

CENTER-RELATED ACTIVITY

Policy Variables		Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	-.003	1.13	.26	-.13	.02
	Child-related education/ training	-.001	.03	.91	-.03	(.04)
II	Observed staff/child ratio	.832	3.96	.000	.41	.16
	Child-related education/ training	-0.12	.49	.65	-.03	(.18)
III	Observed group size	-.003	1.00	.32	-.13	.02
	Child-related education/ training	-.004	.14	.89	-.03	(.04)
	Experience in current day care center	.001	.10	.91	.02	

Table 3.17

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

ATTENTION TO STAFF

Policy Variables		Ordinary Least Squared Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	-.001	.46	.65	-.04	.00
	Child-related education/ training	-.002	.10	.90	-.07	(.08)
II	Observed staff/child ratio	.427	3.83	.000	.38	.15
	Child-related education/ training	-.008	.66	.51	-.07	(.21)
III	Observed group size	-.001	.41	.68	-.04	.01
	Child-related education/ training	-.003	.22	.82	-.07	(.09)
	Experience in current day care center	.001	.57	.57	.38	

Table 3.18

RESULTS OF REGRESSIONS OF
CAREGIVER BEHAVIOR VARIABLES, SPRING 1977
(Lead Teachers, n=87)

ADULT-RELATED ACTIVITY

Policy Variables		Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
I	Observed group size	-.002	1.32	.19	-.15	.05
	Child-related education/ training	-.018	1.56	.12	-.17	(.12)
II	Observed staff/child ratio	.022	.17	.87	-.02	.04
	Child-related education/ training	-.029	1.90	.06	-.19	(.10)
III	Observed group size	-.002	1.43	.16	-.15	.06
	Child-related education/ training	-.027	1.40	.11	-.17	(.13)
	Experience in current day care center	.000	.63	.55	-.03	

in preparation activities--and more time with other staff. This finding suggests that higher ratios give teachers more opportunity for "time out" during the day but also decrease the total amount of lead teacher time available to children. High ratios of course do not necessarily decrease the amount of adult time available to children, since aides may make up the difference. (Chapter Four explores this issue from the child's viewpoint.)

Consistency of Group Size and Ratio Effect: Other Samples

The foregoing results for lead teachers in the 49-center sample show that group size and ratio were related to many of the same teacher behaviors in a pattern suggesting that larger classrooms and low ratios were disadvantageous. Larger group sizes and lower staff-child ratios were associated with more management behavior and more observing; also teachers in larger classrooms and those with lower ratios spent more time with groups of 13 or more children and less time with smaller groups.

A similar pattern of effects was revealed in all of the other samples. There were no contradictions across the samples, and many effects were consistently significant.

- For 49-center aides, the findings for the group composition variables were highly consistent with the teacher findings. There was no evidence of an interaction between caregiver role in the classroom and the policy variables.
- In the APS samples, the findings for group size were not only consistent with the 49-center findings but also stronger. This was especially true for APS aides, for whom there was a significant effect for group size on virtually every dependent measure, with larger groups associated with less caregiver/child interactions of all types, more OBSERVES and ATTENTION TO ADULTS, and more ATTENTION TO SMALL GROUPS. The ratio effects for APS and 49-center samples also were convergent, although there were fewer effects for ratio in the APS sample, compared with 49-center samples.

- The fall 49-center sample revealed fewer significant effects for the group composition variables. Most of the significant effects in the fall samples were significant in the spring samples, although the reverse was not true. The TO WHOM codes showed more consistent effects than the WHAT codes and constructs.
- In the stratified subsamples of the spring 49-center teacher data, the significant effects for the group composition variables were scattered, at least in part because of small sample sizes. Again, there were no contradictions but only on a few variables were effects significant in all subsamples. This was more often true for ratio and for the TO WHOM codes and OBSERVES.

Caregiver Qualifications

As noted earlier, four caregiver qualifications variables were initially tested in the regressions. In the 49-center lead teacher sample only two--specialization and experience in current center--had effects. The other two--years of education and previous day care experience--had no effects in the regressions and so are not discussed below.*

*There are two important points to note regarding the absence of effects for these variables. First, although years of education has no effects in the regressions, at the level of simple correlations its effects were similar to those for specialization. That is, caregivers with more education tended to do more social interacting with children. The problem in assessing effects for years of education was its confounding with the covariables, caregiver race and classroom SES. The confounding meant, first, that when years of education was entered in the regressions along with the covariables, education was never significantly related to lead teacher behavior. Second, the confounding meant that the simple correlations for years of education could not be interpreted as simply education effect but as a complex of variables including education, race and SES. Second, the measure of experience in current context was confounded with ratio (r=.41). Therefore, the effects for experience, which are consistent with effects for ratio, cannot be disentangled from ratio effects.

Also, because of the relatively narrow range in the lead teachers' previous day care experience , the variable was transformed into a binary variable, with a value of "1" for some experience, regardless of amount. Comparison of teachers with some and no experience showed no significant differences. As was true for the continuous variable, the transformed binary version of PREVIOUS DAY CARE EXPERIENCE did not have any relationships to caregiver behavior.

SOCIAL INTERACTION (Table 3.9). Whether teachers had child-related education/training was strongly related to the amount of social interaction with children. Teachers with specialized preparation tended to do more social interacting. This was true especially for the "warm" behaviors, (praise, comfort and respond). Experience in current center was not significantly related to SOCIAL INTERACTION.

MANAGEMENT (Table 3.10). Neither child-related education/training nor experience in current center was related to the amount of managing a teacher did.

OBSERVES (Table 3.11). None of the measures of caregiver qualifications were associated with the amount of observing a teacher did.

ATTENTION TO ONE CHILD; SMALL, MEDIUM and LARGE GROUPS (Tables 3.12-3.15). How a teacher distributed her time was not strongly related to specialized training.

Non-Child Activities: CENTER-RELATED ACTIVITY ADULT-RELATED ACTIVITY and ATTENTION TO ADULTS (Tables 3.16-3.18). Caregivers with child-related education/training tended to spend less time in adult-related activity than those without such training. Relationships beetwen specialization and other "non-child" activities were rather uniformly negative but nonsignificant. A teacher's experience in the current center was not related to the amount of non-child activities.

Consistency of Effects for Caregiver Qualifications in Other Samples

Unlike the group composition variables, the qualifications variables did not have consistent effects across comparison samples--49-center aides, fall data for

teachers in the 49-center study, or stratified subsamples. There appear to be two primary reasons for the inconsistency: (1) The qualifications variables had very different distributions and intercorrelations in the various comparison groups; and (2) as indicated earlier, sample sizes and statistical power were diminished for the comparison groups.

- For 49-center aides, only experience in current center had a "significant" effect, with more experience associated with less observing. Specialization, years of education, or previous day care experience never reached a conventional level of significance ($p<.05$) for aides. One reason for the null finding regarding specialization, which had been a significant factor in the behavior of lead teachers, may be that very few aides used training or education related to young children.

- In the fall 49-center samples, only specialization was tested. In the fall data, it was associated only with more management. Thus, there was no consistency with the spring data, although no contradictions as well.

- The stratified subsamples of the spring 49-center teacher data also revealed no consistent effects for specialization, in part because of statistical power.

Covariables

The covariables (race of caregiver and socioeconomic status of the children) had as strong an effect on lead teacher behavior as any of the policy variables. Both covariables were entered in each of the regression models, and they added significantly to the prediction of caregiver behavior. Since the two were highly correlated ($r = .56$), usually only one was significant for any dependent measure,

and most often it was race of caregiver. White teachers, who were in classrooms with higher average SES, tended to do more social interacting and less observing and adult-related activity. They also tended to spend more time with individual children, less time with medium-sized groups, and more time with other staff.

It is important to note that the covariables did not alter the relationship of the policy variables to the dependent measures. That is, although there were many significant relations between the covariables and caregiver behavior, there was virtually no interaction of the covariables and policy variables. In the tables of regression results, only the coefficients for the policy variables are given, but the R^2 accounted for by the covariables is indicated for each of the dependent measures. It is clear that the covariables frequently were responsible for much of the variance explained.

Consistency of Effects for Covariables in Other Samples

The effects of the covariables in the other relevant 49-center samples (fall samples, spring aides) were always consistent in trend although varied in strength.

- For 49-center aides, there were fewer effects for the covariables. Class SES never was a significant predictor, while caregiver race had a significant effect for RESPONDS, similar in direction to that for teachers.

- For the fall 49-center samples, the effects for covariables were slightly stronger than in the spring data. In the fall teacher data only, management was related to caregiver race (more among white caregivers).

Summary

The analyses of the AFI data showed that caregiver behavior was indeed related to regulatable aspects of the classroom environment. Behavior was strongly related to group composition measures and related to caregiver specialization; there was some evidence of association with amount of caregiver experience.

Group Size and Ratio

AFI results highlight the need to think of group size and staff/child ratio as facets of a larger construct--group composition. Small groups and high ratios had overlapping (though not identical) effects. Both group size and ratio were associated with what caregivers did and with distribution of the caregiver's attention. The effects were consistent across the data samples and stood up in tests of the robustness of the effects (i.e., in biweighted regressions). Regarding codes representing attention to children, ratio and group size were associated with the same measures in opposite directions. Caregivers paid more attention to large groups and less to small groups in larger classrooms; the opposite pattern was associated with higher ratios. Regarding what caregivers did, higher ratios and smaller classrooms were associated with less management and less observing. (Table 3.19).

Though we stress the interrelatedness of group size and ratio, we also attempted to separate their effects, insofar as this was possible, given their degree of correlation. Other NDCS analyses (of child behavior and test performance) suggest that group size is a more powerful predictor than ratio,* and this general finding was confirmed to some degree in the AFI data. When both measures were entered

*See Chapters 4 and 5 in this volume, and relavant chapters in Volume IV of the NDCS Final Report.[6,7]

Table 3.19

SUMMARY OF SIGNIFICANT REGRESSION RESULTS* FOR SPRING AFI, LEAD TEACHERS

	Social Interaction	Manage	Observe	Center Activity	Adult Activity	Attn. to Staff	Attn. to Child	Attn. to Small Group	Attn. to Med. Group	Attn. to Lg. Group
Group Size		(+)	+					-	-	+
Staff/Child Ratio		-	-	+		+		+	+	-
Specialization	+				(-)				(-)	
Experience in Current Center								(+)	-	
Years of Education										
Previous Day Care Experience										
CG Race/Class SES (+=white cg, higher SES)	+		-		(-)	+	+		-	

*Results noted were significant at p<.05,; results significant at .05<p<.15 shown in parenthesis.

in the same regression, the group size effects were strengthened, to the point of reaching significance for ATTENTION TO INDIVIDUAL CHILD, and the ratio effects slightly diminished (Table 3.20). Thus there is some evidence that with ratio controlled, effects for group size not only hold up but are strengthened, whereas effects of ratio are less powerful with group size controlled.

In addition, though ratio and group size had many effects in common, ratio also had some unique effects. It was related to time in non-child activities: caregivers, particularly teachers, in higher-ratio classrooms spent more time in center-related activities and in interaction with other staff. These findings may imply that high ratios put teachers into a more managerial role with the other staff, most of whom are likely to be aides.

Caregiver Qualifications

The single most important finding regarding caregiver qualifications was the relation of specialization to lead teacher behavior. Lead teachers with specialization engaged in more social interaction with children and less non-chld activity than those lacking such preparation. (Table 3.19.) Effects of specialization observed for lead teachers in the 49-center study were not confirmed in other samples, although they were generally not contradicted. The lack of confirmation may have been due in some cases, to inadequate statistical power to detect effects, and in the specific case of the aide sampple, to the fact that few aides have training or education related to young children.

Beyond the effects of specialization, effects of other qualifications variables were few and inconsistent. There was scattered evidence suggesting that the caregiver's

Table 3.20

SUMMARY OF SIGNIFICANT REGRESSION RESULTS* FOR GROUP COMPOSITION MEASURES IN DIFFERENT REGRESSION MODELS, SPRING AFI, LEAD TEACHERS

		Social Interaction	Manage	Observe	Center Activity	Adult Activity	Attn. to Staff	Attn. to Child	Attn. to Small Group	Attn. to Med. Group	Attn. to Lg. Group
Group Size alone			(+)	+					-	-	+
Ratio alone			-	-	+		+		+	+	-
Group Size)	in combination	(-)	+	+		(-)		-	(-)	-	+
Ratio)			-				+	(+)	(+)	-	

*Results noted were significant at $p<.05$,; results significant at $.05<p<.15$ shown in parenthesis.

experience in her current center is associated with positive behaviors; however, as will be seen, there was little support for this suggestion in data based on children's behavior or test scores. Formal education appeared, in simple correlations, to have some effects, but these proved to be bound up with the race and SES composition of the class, and they did not hold up in regression analyses.

CHAPTER FOUR: THE CHILD--BEHAVIOR IN THE CENTER*

BACKGROUND

Behavior of young children in day care is varied and volatile--much more so, for example, than behavior of children in elementary school settings. The NDCS required an observation instrument and analytic approach that could do justice to this complexity, yet yield a manageable set of behavior descriptors that reliably characterized children, classes or centers along dimensions relevant for assessing quality of care.

The study's initial approach used naturalistic observations in combination with standardized tests and rating scales to measure selected characteristics of individual children--traits, dispositions, skills and knowledge--which were potentially susceptible to change due to the child's day care experience. However, several of the standardized tests and rating scales proved to be psychometrically unsound; only two measures of school-related cognitive and linguistic skills--the Preschool Inventory and the Peabody Picture Vocabulary Test--were adequate to support the change score analyses envisioned in the original study design. (See Chapter Five.)

Observations of children's behavior also failed to yield usable trait measures. As indicated in Chapter Two, observation measures were not reliable at the child level. Only when averaged to class level were they moderately reliable, and even at class level they would not support change score analysis. Moreover, though built upon individual scores, these aggregated measures could not

---

*This chapter is based largely on work by David Connell, reported in greater detail in Volume IV of the NDCS final report.[1] Dr. Connell is co-author of this chapter, along with Dr. Jeffrey Travers.

be interpreted simply as class averages of individual traits. Rather, they reflected a blend of individual characteristics and classroom dynamics. There was no evidence to indicate to what degree patterns of child behavior captured by the observation measures would generalize to settings outside of day care or last beyond the preschool years. This situation was not merely a limitation of NDCS instruments and measures; it was but one manifestation of the general difficulty of finding trait measures for young children that show either cross-situational generality or longitudinal stability.

However, whatever their shortcomings as trait measures, observations revealed a great deal about the day-to-day experience and behavior of the child. They were extremely useful in describing the social environment of the day care center and assessing its relationship to regulatable center characteristics. In a sense they provided NDCS researchers with some of the indicators of quality that are available to parents in choosing a day care center for their child--impressions of the degree to which the center provides stimulating social interaction among children and between adults and children, and elicits cooperative, creative and verbal/intellectual activity on the part of the child.

The Child-Focus Instrument

The Child-Focus Instrument (CFI), used in the NDCS for naturalistic observation of children, was based on the Child Observation System developed by Elizabeth Prescott.[2] SRI selected the Prescott instrument after reviewing several alternative systems and conducting field tests of the most promising candidate instruments during Phase I.[3] The Prescott instrument was attractive because it had been developed specifically for preschool children

in day care settings and because it had been used for research purposes quite similar to those of the NDCS. The system includes a large number of behavior codes, many of which are highly specific and have a fairly high degree of face validity and objectivity. SRI was able to train observers to acceptably high levels of accuracy for almost all codes, both in initial field-testing and in subsequent use during Phase II and III (see below).

The CFI was modified several times in the course of the NDCS; the version described here is the one used in Phase III. Each child observation consists of a twenty-minute period, broken into 100 twelve-second coding intervals. Observers are provided with timers that click every twelve seconds. Observers are instructed to record the behavior of a preselected focus child at the time of each click. Each record or frame has three parts.

- A section containing one of 50 codes characterizing the child's principal behavior during the 12-second coding interval. These include 37 activity codes, used when the child engages in some form of overt action, and 13 "receives" codes, used when the major event during the coding interval is an initiative directed toward the child by some other person, e.g., a request, praise, or correction. Additional codes accompany some of the "receives" codes to indicate whether the child's response is appropriate.

- A section containing one of four object codes (adult, child, group of children, or environment), indicating the person(s) or thing(s) toward which the focus child's attention is directed.

- A section containing one of three activity continuity codes, indicating whether the child's behavior is a new activity, an old activity, or no identifiable activity at all.

Table 4.1 lists the codes and shows their relative frequencies of occurrence in the Phase III data; that is, their frequencies as percentages of all 725,000 frames recorded in fall and spring.* Definitions of the more important codes are provided immediately below. Descriptions of the data base and data-gathering procedures appear in the following section.

Many of the CFI codes shown in Table 4.1 are specific and self-explanatory. However, some of the most frequently occurring codes (e.g., "shows closed, structured activity") are broader and require some explication. The following definitions of the most common activity and "receives" codes have been excerpted from SRI's training manual:[4]

> Participates in group activity--closed, structured: Focus child and others are involved in an activity that has a goal, clear guidelines for carrying out the task, and a defined beginning and end. Focus child's participation in adult-directed group activities is coded here. (The presence of other children in the activity differentiates this code from individual structured activity, discussed below.) Examples: child is part of a group playing musical chairs; or child and a friend are working together to clean off the table.

*Frequencies of the activity continuity codes indicating old vs. new activities are not shown directly in the table. By a procedure outlined in the later section on construction of dependent variables, these two codes were used to compute the duration of the child's longest single activity during the 20-minute observation period. The latter figure is shown in the table.

Table 4.1

FREQUENCIES OF CHILD OBSERVATION CODES[a]
(FALL, 1976 AND SPRING, 1977)

A. Activity Codes	Percent of All Frames
Group closed, structured activity	21.1
Group open, expressive activity	13.2
Monitors environment (looks, watches)	11.9
Gives opinions	8.0
Wanders aimlessly, does nothing	5.3
Group passive behavior	4.8
Moves with purpose	3.1
Individual open, expressive activity	2.9
Adds prop or idea	2.8
Considers, contemplates, tinkers	1.7
Individual closed, structured activity	1.5
Gives orders, directs others	1.0
Intrudes playfully	0.9
Asks for attention	0.9
Selects activity (with others)	0.6
Shares, helps	0.6
Asks for information	0.4
Asks for turn	0.3
Selects activity (alone)	0.3
Isolates self	0.3
Asserts rights	0.3
Cries	0.2
Sees pattern, solves problem	0.2
Intrudes hostilely, bullies	0.1
Hostilely asserts rights, anger	0.1
Hostile exchange	0.1
Avoids, withdraws	0.1
Individual passive activity	0.1
Asks for assistance, help	0.1
Offers sympathy, comfort	0.1
Asks for comfort	0.1
Intrudes unintentionally	0.1
Experiences rejection	0.1
Quits activity after frustration	<0.1
Angry reaction to frustration	<0.1
Experiences accident	<0.1
Temper tantrum	<0.1

Table 4.1 (continued)

B. "Receives" Codes[b]	Percent of All Frames	Percent of Appropriate Responses
Receives general comments	5.1	86.5
Receives information, guidance	4.7	87.1
Receives demands, requests	4.0	82.3
Receives request to play, share	0.5	63.1
Receives rules, corrections	0.4	70.4
Receives punishment, threats	0.3	47.7
Receives praise	0.3	82.6
Receives playful intrusion	0.9	
Receives comfort	0 2	
Receives hostile intrusion	0.1	
Receives unintentional intrusion	0.1	
Receives physical punishment	<0.1	
Receives rejection	<0.1	
C. Object Codes		
Attention to adult	27.3	
Attention to child	23.0	
Attention to group	7.8	
Attention to environment	41.9	
D. Activity Continuity Codes		
Longest Activity[c]	54.8	(11 Minutes)
Not involved in activity	7.3	

[a]Code frequencies are shown as a percentage of all observation codes (excluding structured situation observations). For both behavior and object codes, the total number recorded was approximately 725,000.

[b]The "receives" codes indicate initiatives by others toward the child. The column headed "Percent of All Frames" shows the frequencies of these codes as percentages of all 725,000 codes. The column headed "Percent of Appropriate Responses" indicates how often children responded appropriately to selected initiatives.

[c]The "longest activity" code is computed as a percentage ratio of the duration of the longest activity to the total duration of the observation period. Since the observation period usually lasted 20 minutes, the longest activity of the typical child lasted .548 x 20 minutes, or 11 minutes.

Participates in group activity--open-ended, expressive: Focus child participates with others in a mutual experience that has no goal, no external guidelines or defined point of completion; the structure of the activity is determined by those involved, not by the materials. (The presence of other children in the activity differentiates this code from individual open-ended activity, discussed below.) Examples: Child is playing with other children in the block corner; or child and another child are swinging alongside each other, making a game of who can swing higher.

Monitors environment (looks, watches): Focus child's attention is obviously directed at other people or things. This code is not used for listening. The focus child may be either in or out of an activity. The Object code used with this code identifies the focus of the child's attention. Examples: Child stands apart from a group of children, watching them play; or child is playing at the block table, and his attention is directed to an adult cleaning up some spilled paint.

Gives opinions, preferences, information, comments: Focus child initiates statements about his own likes, dislikes, or preferences. This code also includes information and comments initiated by the focus child (not in response to a question). Examples: "I went on a picnic yesterday"; or "Johnny is my best friend."

Does nothing, wanders: Focus child wanders around center with no apparent purpose to his movement. He may be sitting or standing doing nothing, looking around the area with no apparent focus. Examples: Child wanders from sandbox to slide and then to doll corner, not concentrating on anything or anyone.

Participates in group activity--passive attention: Focus child is part of a group that is involved in an activity which requires no visible response, but does require concentration or thought. (The presence of other children in

the activity differentiates this code from monitoring the environment.) Examples: Child and other children are watching a puppet show; or child is part of a group that is watching TV; or child is part of a group to which an adult is reading a story.

Moves with purpose: This code is used when the focus child is going from one activity to another or whenever it seems evident that there is some goal to his movement. Examples: Child has just finished gluing on a piece of paper; he heads for the bathroom to wash his sticky hands; or, child notices that a swing is free and runs across the yard toward it.

Individual open-ended, expressive activity: Focus child is involved in an activity that has no defined goal, external guidelines, or defined point of completion; the structure of the activity is determined by the child. Other children do not share in this activity with the focus child--he is alone. Examples: Child is playing with blocks; or child is dancing alone to a record.

Adds a different prop or new idea: Focus child adds variety to his activity. He uses a different toy or prop from the one he was using previously in the same activity, or he uses the same prop in a different way. This code is also used when the focus child resumes play with an article that he used formerly in the same activity. Examples: child adds a different color to his painting; or child is washing dishes in the doll corner, then picks up a doll and washes it.

Considers, contemplates, tinkers: Focus child considers before making a selection of materials. Focus child tries out an object, looks at it, moves it, examines it, manipulates it. Focus child struggles with a problem, attempting to solve it. Examples: Child carefully examines a truck, checking out each moving part; or child pulls on cargo net and watches how the net moves in response to his pull.

Individual structured, closed activity: Focus child is involved in an activity which has a goal, clear guidelines. Other children do not share in this activity with the focus child. Examples: Child is stringing beads for a necklace; or child is working on a puzzle; or child is alone at a table, grating cheese for a pizza.

Receives orders or minor behavioral corrections: Focus child receives commands with which compliance is expected. This code also includes orders to maintain smooth operation of the center and minor behavioral corrections. Examples: Adult tells child to put books away; or another child says to focus child, "Let me have the trike now."

Receives information/help with a task: Focus child receives instruction, materials, or assistance related to his task or the solution to his problem. This code includes verbal and nonverbal assistance or demonstration. Also included in this code are preliminary directions and review of an activity. Examples: Child is having difficulty completing his puzzle and the teacher shows him where the piece goes; or adult is telling focus child how to clean paint brushes.

Receives general comments, questions: Focus child is asked for information or receives comments of a general nature. Examples: Adult says to child, "Today is Johnny's birthday"; or another child tells focus child, "My grandma made this dress."

Frequencies of the behavior codes varied widely in Phases II and III. In Phase III, all of the eleven activity codes and three "receives" codes defined above occurred more than once per 20-minute observation (i.e., more than one percent of the time). Most analyses reported in later sections are based on these common codes and combinations thereof. However, many codes of psychological interest occurred rarely--a few times per thousand frames, or less.

Many of the latter were events that are potentially important as indicators of harm; a few were potential indicators of benefits of day care. Examples include the codes "cries," "isolates self," "refuses to comply," "experiences accident," "shares or helps," and a number of codes indicating anger or hostility.

There are several possible reasons for the low frequencies of these events. One, mentioned in Chapter Two, is that frequencies of events recorded with a time-sample instrument such as the CFI depend partially on the durations of those events. If psychologically important events are brief, they will appear in few frames or be missed altogether. A second reason has to do with limited opportunities for children to display behaviors that meet the definitions of relevant codes. For example, sharing, taking turns and helping with minor tasks are routinized in most centers. Routinized prosocial behavior is coded as a form of group activity, or as compliance with adult requests, rather than as voluntary helping or sharing, accounting for the rarity of this particular code. Similarly, most centers are organized to prevent conflict and to terminate it quickly when it occurs. To the degree that they succeed, "opportunities" for conflict are limited, and associated codes are rare.

Two approaches were taken in dealing with the rarity of important codes. First, in addition to natural classroom observations, children were observed in structured situations designed to provide greater opportunity for voluntary prosocial behavior such as helping and sharing. Second, rare codes from the natural observations were analyzed separately from more frequent codes, using a form of statistical analysis more appropriate for rare events than ordinary regression. Results of both approaches are

presented in separate sections at the end of this chapter, following discussion of the main analyses and results.

Phase III Sample and Procedures

The study design called for each child to be observed four times for a total of eighty minutes in both fall and spring--three times in natural situations (primarily free play and teacher-directed activity) and once in a pair of structured situations. In the spring, natural observations were conducted by two different observers for each child--generally one black observer and one white --in order to permit analytic separation of actual behavioral differences among children from differences in perspective among observers. SRI was able to implement the design with substantial success, as the following data indicate.

Approximately 8,300 twenty-minute observations of target children were completed by SRI's observers. The distribution of observations between time points and between natural (classroom) and structured observations is shown in Table 4.2. Numbers of children and classrooms observed are also shown in the table. Of 1,108 children observed in the spring, 1,086 had been observed in the fall. At both times, the sample was approximately evenly divided among Atlanta Public School centers, Atlanta centers outside the public schools, Detroit centers and Seattle centers.

In both fall and spring, natural observations took place in four general types of situations: free play; adult-directed activity (including both individual and group activities, with the latter predominant), routine center activities (cleanup, snack, toileting, etc.) and multiple activities--combinations of two or more of the preceding

Table 4.2

PHASE III CHILD OBSERVATION SAMPLE

	Fall 1976	Spring 1977
Natural (Classroom) Observations		
Number of Observations	3,987	3,177
Number of Children	1,310	1,108
Number of Classrooms	117	116
Structured Observations		
Number of Observations	642	523
Number of Children	1,284	1,046

types occuring within one twenty-minute observation. By design, free play and teacher-directed activities were observed most frequently. About 38 percent of fall observations and 41 percent of spring observations took place during free play periods; 42 percent of fall observations and 41 percent of spring observations occurred during teacher-directed activities. Since the dynamics of the group can change dramatically across these general types of situations, separate analyses were conducted for data from free play and teacher-directed periods. In addition, selected analyses were performed on data pooled across all four situations.

SRI hired and trained 46 observers in both fall and spring. Each time, nine observers conducted structured observations exclusively, while the remaining 37 conducted natural observations in classrooms. Between fall and spring, the number of observers who were members of minority groups was increased from 12 to 20, or 44 percent of the total. These observers completed 44 percent of all observations, close to the 50 percent ideally required by the study procedures discussed in Chapter Two. A minimum of 30 percent of observations in each center were conducted by minority observers. All observers were female. Distributions of age and education were fairly similar across sites; most observers were college graduates between 30-35 years of age.

DESCRIPTION OF VARIABLES

Selection and Construction of Dependent Measures

With child observations as with observations of adults, the study's general strategy was to describe behavior in the day care center as comprehensively and objectively as possible, in terms of fine-grained codes. Data were then

reduced by combining frequencies of codes that were conceptually related and empirically correlated. Efforts were made to create summary variables that bore some relationship to constructs previously used in the developmental literature, but primary weight was placed on empirical patterns evident in the data. As with the adult observations, relatively little data reduction proved to be appropriate. The dependent variables ultimately used in exploring relationships between regulatable center characteristics and child behavior were a mix of individual codes and a few summary measures.

In one effort to reduce the set of codes to a few summary dimensions, principal components analyses were performed on child- and class-level data from the fall and spring samples. The principal components analysis proved unrevealing. The resulting dimensions accounted for little variance and were not readily interpretable. Nor were they especially stable from fall to spring. Moreover, some "dimensions" were dominated by one or two particularly frequent codes. Consequently, conceptual coherence and simple correlations among codes were the primary bases for deciding how to combine codes to form broader constructs.

To choose appropriate combinations of codes, frequencies and correlations among various codes were examined, at all levels of aggregation--child, class and center. Data were also examined separately for fall and spring, for the Atlanta Public School classrooms, and for the three sites of the 49-center study. This approach led to identification of a number of candidate measures, of which four are discussed in this report. Of the four, two--called REFLECTION/INNOVATION and COOPERATION/COMPLIANCE--proved to be related to the policy variables. Two others--INTEREST/PARTICIPATION and the CLASSROOM ACTIVITY BALANCE are also discussed here because of their descriptive

interest and because the latter proved to be related to children's test performance. (See Chapter Six.) All four variables are defined in the next section.

In addition, eight dependent measures based on individual codes are discussed here. The eight codes are singled out because they were relatively frequent, distinct in meaning from other codes, related to the policy variables and collectively were judged to reflect some important aspects of this quality of care. Four were codes denoting the object of the child's attention--ORIENTATION TO ADULTS, INDIVIDUAL CHILDREN, GROUPS OF CHILDREN and THE ENVIRONMENT--which describe the child's global interaction patterns. The remaining four were the code "gives opinions, etc." (VERBAL INITIATIVE), longest activity (TASK PERSISTENCE), "does nothing, wanders" (AIMLESS WANDERING) and the continuity code "no task" (NONINVOLVEMENT). (Again, as indicated earlier, some infrequent codes representing psychologically important events were treated differently and are discussed separately.)

Along with definitions of the various measures, the next section contains information on the consistency of each measure across adult-directed and free play activity periods (indicating the degree to which the measures characterize classrooms rather than activity segments within classrooms). Age trends are also reported when important, and selected correlations among the measures are reported wherever these help clarify the meaning of a particular measure. Finally, stabilities of measures from fall 1976 to spring 1977 are also reported. Stability correlations identify those constructs for which center classrooms retain their relative frequency rankings from fall to

spring, as opposed to those constructs for which classrooms shift noticeably in relative frequency ranks. These measures give some indication of which behavior patterns are established rapidly during the day care year* and which patterns take shape gradually from fall to spring. However, the correlations are somewhat underestimated because of changes in observation procedures from fall to spring discussed in Chapter Two and because of shifts of enrollment within classes.**

REFLECTION/INNOVATION

Two codes--considers, contemplates or tinkers and adds prop or idea--came closest among all CFI codes to capturing thoughtful, creative problem-solving behavior on the part of children. Because of their low individual frequencies and positive correlations (.34 in fall, .30 in spring), the two were summed to form a statistically more robust variable, REFLECTION/INNOVATION. Frequencies of the construct tended to be consistent across activity periods ($r=.42$, $p<.01$ in fall; $r=.37$, $p<.01$ in spring) but were unstable from fall to spring.

---

*The "day care year" is not as sharply defined as the school year, with a clear beginning in fall and in spring. However, formal and anecdotal NDCS data from both the Supply Study and main cost-effects study show that there is a major influx of new children in the fall, accompanied by an exodus of children who have reached school age. There is also a drop off of enrollment during the summer months.

**Correlations of code frequencies between free play and teacher-directed activities are based on 117 classrooms in fall and 116 in spring. Fall-spring stability correlations are based on 114 classrooms that existed at both time points, although shifts in enrollment occurred within those classrooms.

VERBAL INITIATIVE

The single code gives opinions, preferences, information, comments was treated as a separate variable indicating the degree of verbal self-assertiveness exhibited by children and expected or accepted by caregivers. Frequencies of VERBAL INITIATIVE were consistent across activity types ($r=.62$, $p<.01$ in fall; $r=.32$, $p<.01$ in spring) but had only modest fall-to-spring stability ($r=.18$, $p<.05$ for free play; $r=.12$, n.s. for adult-directed activity).

COOPERATION/COMPLIANCE

Seven of the "receives" codes are accompanied by supplementary codes indicating whether the child's response is appropriate. The seven relevant categories of action or statement directed toward the child are (1) general comments, (2) information or guidance, (3) requests to play or share, (4) demands or requests other than requests to play or share, (5) rules or corrections, (6) punishment or threats, and (7) praise. Percentages of appropriate responses, shown in Table 4.1, ranged from a low of 48 percent for punishment and threats to 87 percent for comments, information and guidance. An index of COOPERATION/COMPLIANCE was computed as the ratio of all active appropriate responses to all instances of these seven "receives" codes. In the fall, older children showed higher frequencies of COOPERATION/COMPLIANCE than younger children ($p<.05$), but no age differences were evident in spring--perhaps indicating a progressive socializing effect for younger children. Cooperation was at best marginally consistent across activity periods ($r=.18$, $p<.05$ in fall; $r=.08$, n.s., in spring). Cooperation during free play was moderately stable from fall to spring ($r=.25$, $p<.01$) but cooperation during adult-directed activity was not ($r=.06$, n.s.).

NONINVOLVEMENT

The degree to which children are uninvolved in classroom activities is directly recorded by the activity continuity code no task ("Task" is broadly defined and includes play and exploration as well as teacher-assigned activities). This index of NONINVOLVEMENT was consistent across activity types ($r=.50$, $p<.01$ in fall; $r=.34$, $p<.01$ in spring) and was stable from fall to spring for adult-directed activity ($r=.44$, $p<.01$), but much less so for free play ($r=.11$, n.s.).

AIMLESS WANDERING

Like NONINVOLVEMENT, AIMLESS WANDERING--measured by the frequency of the code does nothing, wanders--is an index of the degree to which children are not engaged in classroom activities. The two variables are correlated ($r=.28$, $p<.01$, for free play, and $r=.45$, $p<.01$, for teacher-directed activity). However, the two were not summed to form a single construct because they were incommensurate. Does nothing, wanders was an activity code, one of 50 possible, whereas no task was a continuity code, one of three possible. No task was often recorded along with does nothing, wanders, accounting in part for their correlation and rendering their sum meaningless. The frequency of AIMLESS WANDERING was consistent across activity types ($r=.40$, $p<.01$ in the fall, and $r=.52$, $p<.01$, in the spring) and was moderately stable from fall to spring ($r=.28$, $p<.01$ for all activity types pooled).

TASK PERSISTENCE

The concepts "task persistence" and "attention span" commonly refer to a child's tendency or ability to

devote sustained effort to a single pursuit. Increasing the young child's capacity in this area is often regarded as an important goal of early education. The focus here is less on task persistence and attention span as individual traits than on closely related characteristics of the classroom, namely demands made and opportunities provided for sustained activity. The CFI provides an indirect measure of these constructs. The activity continuity code designated old activity marks every occasion on which a child continues an activity from one twelve-second interval to the next. By summing durations of all intervals so marked, between the outset of the activity (indicated by a new activity code) and its termination (indicated by another new activity code or a no activity code) it is possible to measure the total duration of every activity in the twenty-minute observation period to the nearest twelve seconds. The mean duration of each child's longest activity, shown in Table 4.1, is 11 minutes. Phase III data, consistent with Phase II findings and previous research, show that activities last longer, on the average, in groups of older children than in younger groups. Moreover, activities last longer in groups where structured activities predominate. The correlation between activity length and the "classroom activity balance" (defined below) was -.37 ($p<.01$) in fall and -.48 ($p<.01$) in spring. However, longest activity was neither strikingly consistent across activity types nor stable from fall to spring.

ORIENTATION TO ADULTS

ORIENTATION TO ADULTS was, predictably, twice as frequent in caregiver-directed activity as in free play. However, frequencies showed fairly high correlations across the two types of activity period ($r=.43$, $p<.01$ in fall, $r=.36$, $p<.01$ in spring), indicating that some groups of children were consistently more adult-centered than others, regardless of prevailing activities. The construct was more stable from fall to spring for free play ($r=.43$, $p<.01$) than for adult-directed activity periods ($r=.08$, n.s.).

ORIENTATION TO INDIVIDUAL CHILDREN

ORIENTATION TO INDIVIDUAL CHILDREN also showed substantial correlations between free play and teacher-directed activities ($r=.60$, $p<.01$ in fall; $r=.35$, $p<.01$ in spring), again indicating a consistent focus of some classrooms on child-child interchange. Combined frequencies of this variable across the two types of activity period were moderately stable from fall to spring ($r=.29$, $p<.01$). However, the Atlanta Public School subsample, which showed a particularly high level of ORIENTATION TO INDIVIDUAL CHILDREN in the fall, also showed a drop from fall to spring which was not observed in any of the 49-center study sites.

ORIENTATION TO GROUPS

The object code ORIENTATION TO GROUPS was included as a dependent measure primarily to determine whether children's contact with their peers was affected by classroom composition, specifically whether their attention is directed to group rather than solitary or one-to-one activity as total class size grows. Fall-to-spring correlations for this measure were .38, $p<.01$ for free play and .27, $p<.01$ for teacher-directed activity. Consistency across teacher-directed and free play activities was .24, $p<.05$ in fall and .27, $p<.01$ in spring.

Other Measures

Additional CFI measures will not be discussed in relation to the policy variables in the sections which follow. Though a number of significant relationships were obtained between individual policy variables and CFI measures across many analyses (including fall and spring, teacher-directed and free play activities, the 49-center and the APS studies), few consistent or coherent patterns

emerged. In essence, unreported relationships may be regarded as null, or at least lacking confirmation from multiple data sources. This implies, of course, that the relationships discussed below are selected from a larger set and that significance levels for individual analyses are inconclusive. Again, as stressed repeatedly, results must be interpreted in the context of the study's findings as a whole.

Though other measures will not be examined in relation to the policy variables, two in particular contribute descriptive information toward a profile of child behavior, and show some informative links to the variables listed above.

INTEREST/PARTICIPATION was a global variable reflecting the degree to which children in a class are actively involved in its social and educational activities. INTEREST/PARTICIPATION was computed as the sum of many codes (group and individual open, expressive activity; considers, contemplates or tinkers; adds prop or idea; acts creatively or solves problem; offers to help or share; defends rights; moves with purpose; selects activity (alone or with others), asks for information; asks permission to share; gives opinions; asks for recognition; gives orders or directs others; intrudes playfully). The construct is related to a behavior cluster that has emerged repeatedly in studies of preschool children in group care settings and that is associated with children's later social adjustment and cognitive achievement.[5] A similar construct also emerged during Phase II of the NDCS. In both Phase II and Phase III, codes comprising the construct were positively correlated with each other and negatively correlated with codes indicating noninvolvement. INTEREST/PARTICIPATION also was positively related to TASK PERSISTENCE ($r=.22$, $p<.05$ in fall; $r=.26$, $p<.01$ in spring). NONINVOLVEMENT showed negative correlations in the .3-.4 range with INTEREST/PARTICIPATION in both

free play and teacher-directed activity periods. Thus, NONINVOLVEMENT and INTEREST/PARTICIPATION together tend to array classrooms along a general dimension indicating the degree to which children are integrated into classroom activities. In spring, high levels of COOPERATION/COMPLIANCE tended to accompany high levels of INTEREST/ PARTICIPATION and low levels of NONINVOLVEMENT. (No significant relations were found in fall.) In short, though the relevant correlations were not strong, INTEREST/ PARTICIPATION was part of a broad cluster of positive dynamics in the classroom.

A second variable that characterizes the global dynamics of the classroom (but is not related to the policy variables) is the CLASSROOM ACTIVITY BALANCE. The most commonly used CFI codes were participates in group activity --closed, structured and participates in group activity-- open-ended, expressive. These two codes represented about one-third of all activities recorded. When individual structured and open-ended activities were pooled with the respective group activity codes, all four together accounted for over 37 percent of the codes recorded. Class-level correlations between frequencies of structured and open-ended activities were negative and substantial in both the fall ($r=-.36$; $p<.01$) and spring ($r=-.63$; $p<.01$), indicating that classrooms tend to be characterized by one type of activity or the other. Note that activities defined as "closed, structured" should not be equated with educational activities. Rather, the codes represent activities that have a clearcut end point or achievable goal, whereas open-ended expressive activities do not. Either type of activity can be educationally or developmentally valuable. Nevertheless the two types of activity codes seem to capture distinctive classroom styles.

The CLASSROOM ACTIVITY BALANCE, designed to locate a given classroom on the structured/open-ended dimension, was constructed by subtracting the sum of frequencies of group and individual structured activities from the sum of frequencies of group and individual open-ended activities. This difference score averaged -.06 in the fall and -.04 in the spring, indicating a slight prevalence of structured over open-ended activity, and very little change with time in the overall balance among Phase III centers. The relative ranking of different classrooms on the unstructured/open-ended dimensions was moderately stable from fall to spring ($r=.36$; $p<.01$). Open-ended activities were more prevalent in classes with younger children.

Reliabilities of the Dependent Measures

Reliabilities of the CFI measures were assessed in a number of ways. First, in SRI's training observers had to reach a criterion of 75 percent correct identifications of a set of 115 videotaped examples of child behaviors, recorded under field conditions and selected by the SRI training team. Scores in criterion testing ranged from 76 to 96 percent across observers, with a mean of 88 percent. After two weeks in the field, 42 observers were retested on a slightly smaller sample of videotaped behaviors. Most observers improved their scores; none scored lower than 80 percent, and mean accuracy was 93 percent. In addition, SRI conducted a field test of inter-rater agreement to address the issue of racial differences in coding patterns that had arisen in Phase II. Seventeen pairs of observers were formed, each with one black and one white member. Each pair coded the activities of the same child for one hour. Interobserver comparisons were possible for 45 activity codes, of which only three showed significant differences in overall frequency between black and white observers.

Training of observers and results of various tests of observers' accuracy are described in more detail in SRI's Phase III report.[6]

As noted in Chapter Two, generalizability computations also were carried out for selected CFI codes. Analyses of the components of variance suggested that while variation of children's behavior from occasion to occasion was predictably large, classroom aggregates were reliable enough to permit comparisons of groups of classrooms that differed along policy-relevant dimensions.

Approach to the CFI Analyses

Two kinds of analyses were used to explore links between the policy variables and child behavior. Regression analyses were used for the important and relatively frequent codes. These analyses are reported first. Rare but important behaviors were analyzed by logit techniques, which are discussed following the main body of regression analyses.

The regression model used to explain variance in child behavior entered six policy variables and two covariables. The six policy variables were observed group size, observed staff/child ratio, caregiver years of education, training or education in a child-related field, experience in day care prior to employment at current center, and experience in current center. All measures, dependent and independent, were averaged to the classroom level. Thus, measures of caregiver qualifications represent averages for the staff (lead teachers and aides) in each classroom. The two covariables entered were average age of children in the class and a class-level measure of socioeconomic status (SES).*

*The variable for SES of the classroom was a construct representing five measures: parent education, family size, family income, number of parents and race of child. The five variables were factor analyzed and a princial components factor score was assigned to each class.

The eight independent measures were confounded to some extent. Some of the confoundings were shared by all other data sets, as indicated in Tables 2.4 and 2.5. There were moderate correlations between group size and staff/child ratio, years of education and both child-related education/training and previous experience. One confounding unique to the CFI was between staff/child ratio and age of child: Higher ratios are found in classes with younger children. These confoundings indicate where there are limitations on interpreting the effect of a variable as an independent effect.

The regression approach was "hierarchical." First, the effects of average age of children were accounted for, to be certain that age-related differences in child behavior were not mistaken for effects of policy variables. Second, the class SES measure was entered. (Preliminary analyses showed few effects of socioeconomic status on child behavior in the relatively homogeneous APS sample. Consequently, the SES covariable was entered only in the 49-center regressions.) Finally, the policy variables were entered as a group, in stepwise fashion.

Discussion of the CFI findings concentrates on data collected in spring 1977. Not only did spring data collection procedures minimize observer effects, but the data themselves are likely to reflect patterns of child behavior that have stabilized over the year. Fall and spring data were treated as replications of a single study, looking for consistency of findings. The discussion of spring results is followed by a brief consideration of significant divergences between the fall and spring data. In addition, the regression results for the APS centers are discussed.

EFFECTS OF THE POLICY VARIABLES

Child Behavior Results in Spring 1977

The spring data were obtained from observations in 116 classrooms. Results of the relevant regressions are shown separately for each of 9 dependent variables and separately for free play activities and adult-directed activities. (See Tables 4.3-4.11 below.) The effects for the covariables, the group composition variables, and caregiver qualifications are discussed separately. (See Table 4.12 below.)

Covariables

The class-level SES measure had relatively few strong effects in the regressions, but relatively many effects at a level of significance around p=.10. The strongest effects were for NONINVOLVEMENT and TASK PERSISTENCE: higher SES classrooms tended to have more NONINVOLVEMENT and less TASK PERSISTENCE in both activity contexts. In addition, higher SES classrooms tended to have more open, unstructured activities and more attention to adults during adult-directed activities. All of these effects were relatively weak. Some of these effects may not be due to SES itself but may be indirectly influenced by the FIDCR, which mandate high staff/child ratios and small groups, and which primarily affect centers serving low SES populations. To the degree that this interpretation is correct, removal of variance associated with SES may lead to underestimation of the effects of the policy variables. Thus the estimates reported below may be viewed as conservative.

Average age of children in the classroom was related to the social orientation of the child. Older children less often attended to adults or to groups during

Table 4.3

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: REFLECTION/INNOVATION
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy$_2$ Variables (R^2 with Covariables)
DURING FREE PLAY ACTIVITIES					
Observed group size	-.001	1.75	.08	-.25**	.13
Observed staff/child ratio	.009	0.14	.90	.11	(.13)
Child-related education/ training	-.008	0.89	.38	.03	
Staff education	.004	1.62	.11	.13	
Previous day care experience	.004	1.28	.20	.12	
Experience in current day care center	.004	2.44	.02	.20*	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	-.001	1.71	.09	-.19*	.08
Observed staff/child ratio	.010	0.12	.90	.12	(.11)
Child-related education/ training	.018	1.91	.06	.28**	
Staff education	.001	0.49	.63	.12	
Previous day care experience	.001	0.30	.76	.10	
Experience in current day care center	.001	0.56	.58	.14	

*p<.05
**p<.01

Table 4.4

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: VERBAL INITIATIVE
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy$_2$ Variables (R^2 with Covariables)
DURING FREE PLAY ACTIVITIES					
Observed group size	-.001	2.12	.04	-.21*	.11
Observed staff/child ratio	.109	1.40	.17	.15	(.16)
Child-related education/ training	.006	0.62	.54	.24*	
Staff education	.004	1.43	.16	.24*	
Previous day care experience	.003	1.01	.32	.01	
Experience in current day care center	-.002	1.40	.16	-.10	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	-.001	1.87	.06	-.19*	.08
Observed staff/child ratio	.047	0.55	.58	.08	(.11)
Child-related education/ training	.003	0.30	.76	-.01	
Staff education	.003	1.06	.29	.19*	
Previous day care experience	-.004	1.22	.22	-.04	
Experience in current day care center	-.002	1.40	.16	-.15	

*p<.05
**p<.01

Table 4.5

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: COOPERATION
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables
DURING FREE PLAY ACTIVITIES					
Observed group size	-.006	2.44	.01	-.24**	.11
Observed staff/child ratio	-.097	0.26	.79	.08	(.13)
Child-related education/ training	.137	2.97	.00	.22*	
Staff education	-.026	1.97	.05	-.08	
Previous day care experience	-.004	0.24	.81	.07	
Experience in current day care center	-.009	1.17	.24	.04	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	-.005	1.87	.06	-.21*	.06
Observed staff/child ratio	.144	0.46	.65	.13	(.07)
Child-related education/ training	.042	1.18	.24	.11	
Staff education	-.003	0.30	.76	.07	
Previous day care experience	.003	0.28	.78	.10	
Experience in current day care center	-.003	0.47	.64	-.02	

*p<.05
**p<.01

Table 4.6

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: NON-INVOLVEMENT
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
DURING FREE PLAY ACTIVITIES					
Observed group size	.003	3.85	.00	.30**	.19
Observed staff/child ratio	.040	0.41	.68	-.18*	(.33)
Child-related education/ training	-.042	3.32	.00	-.34**	
Staff education	.005	1.34	.18	.03	
Previous day care experience	-.006	1.42	.16	-.26**	
Experience in current day care center	.001	0.68	.50	-.13	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	.001	0.87	.39	.17	.07
Observed staff/child ratio	-.151	1.59	.11	-.26**	(.15)
Child-related education/ training	-.012	1.18	.24	-.21*	
Staff education	-.001	0.42	.67	-.05	
Previous day care experience	.002	0.66	.51	-.03	
Experience in current day care center	.000	0.10	.99	.01	

*p<.05
**p<.01

Table 4.7

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: AIMLESS WANDERING
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
DURING FREE PLAY ACTIVITIES					
Observed group size	.002	2.17	.03	.33**	.17
Observed staff/child ratio	-.229	1.82	.07	-.30**	(.17)
Child-related education/ training	-.003	0.25	.80	-.14	
Staff education	-.004	1.01	.32	-.16	
Previous day care experience	-.006	1.21	.23	-.20*	
Experience in current day care center	-.001	0.34	.73	-.06	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	.002	1.51	.13	.21*	.16
Observed staff/child ratio	-.294	2.77	.01	-.31**	(.17)
Child-related education/ training	-.007	0.18	.86	-.14	
Staff education	-.006	1.45	.15	-.16	
Previous day care experience	-.006	1.31	.20	-.20*	
Experience in current day care center	-.001	0.42	.68	-.06	

*p<.05
**p<.01

Table 4.8

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: TASK PERSISTENCE
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
DURING FREE PLAY ACTIVITIES					
Observed group size	-.005	0.34	.73	-.06	.13
Observed staff/child ratio	.323	1.94	.06	.25**	(.20)
Child-related education/ training	.058	2.82	.01	.31**	
Staff education	-.016	2.69	.01	-.15	
Previous day care experience	.014	1.97	.05	.25**	
Experience in current day care center	.001	0.15	.88	.17	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	.005	1.62	.11	.10	.08
Observed staff/child ratio	.298	2.09	.04	.19*	(.13)
Child-related education/ training	.093	2.22	.03	.21*	
Staff education	-.012	1.86	.07	-.09	
Previous day care experience	.002	0.12	.91	.06	
Experience in current day care center	.002	0.22	.83	.09	

*p<.05
**p<.01

Table 4.9

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: ORIENTATION TO ADULTS
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy$_2$ Variables (R^2 with Covariables)
DURING FREE PLAY ACTIVITIES					
Observed group size	-.006	3.11	.00	-.30**	.11
Observed staff/child ratio	.231	0.84	.40	.18*	(.15)
Child-related education/ training	.028	0.80	.43	-.01	
Staff education	-.009	0.94	.35	-.05	
Previous day care experience	-.004	0.31	.76	.07	
Experience in current day care center	-.009	1.41	.16	-.18*	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	-.007	2.78	.01	-.27**	.11
Observed staff/child ratio	-.105	0.32	.75	.08	(.14)
Child-related education/ training	.026	0.74	.46	.02	
Staff education	-.000	0.02	.96	.09	
Previous day care experience	.009	0.78	.44	.13	
Experience in current day care center	-.002	0.34	.74	-.07	

*p<.05
**p<.01

Table 4.10

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: ORIENTATION TO CHILDREN
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy Variables (R^2 with Covariables)
DURING FREE PLAY ACTIVITIES					
Observed group size	.002	1.17	.25	.14	.04
Observed staff/child ratio	-.078	0.41	.68	-.09	(.06)
Child-related education/ training	-.004	0.16	.87	.05	
Staff education	.006	0.96	.34	.09	
Previous day care experience	-.001	0.08	.93	-.03	
Experience in current day care center	.002	0.54	.59	.04	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	.003	1.33	.19	.10	.08
Observed staff/child ratio	.121	0.47	.64	-.01	(.18)
Child-related education/ training	-.002	0.07	.94	.15	
Staff education	.016	1.99	.05	.28**	
Previous day care experience	-.000	0.02	.99	.02	
Experience in current day care center	.001	0.17	.86	.03	

*p<.05
**p<.01

Table 4.11

RESULTS OF REGRESSIONS OF CHILD BEHAVIOR VARIABLES ON SELECTED POLICY VARIABLES
Dependent Variable: ORIENTATION TO GROUPS
Spring, 1977 (n = 116)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2 for Policy$_2$ Variables (R^2 with Covariables)
DURING FREE PLAY ACTIVITIES					
Observed group size	.003	2.86	.00	.28**	.11
Observed staff/child ratio	.172	1.58	.12	.14	(.16)
Child-related education/ training	.008	0.54	.59	.05	
Staff education	-.003	0.90	.37	-.12	
Previous day care experience	.000	0.04	.97	-.01	
Experience in current day care center	-.004	1.40	.16	-.09	
DURING ADULT-DIRECTED ACTIVITIES					
Observed group size	.004	2.43	.02	.29**	.12
Observed staff/child ratio	-.017	0.20	.84	-.04	(.14)
Child-related education/ training	.012	0.68	.53	.01	
Staff education	-.006	1.25	.21	-.14	
Previous day care experience	-.004	0.54	.59	-.07	
Experience in current day care center	-.008	2.07	.03	-.16	

*p<.05
**p<.01

Table 4.12

SUMMARY OF SIGNIFICANT REGRESSION RESULTS FOR SPRING CHILD OBSERVATIONS[1]

	Reflection/ Innovation		Verbal Initiative		Cooperation		Non-involvement		Wandering		Task Persistence		Orientation to Adults		Orientation to Children		Orientation to Groups	
	FP	AD	FP	AD	FP	AD	FP	AD	FP	AD	FP	AD	FP	AD	FP	AD	FP	AD
Group Composition																		
Group Size	(-)	(-)	-	(-)	-	(-)	+		+	(+)		(+)	-	-			+	+
Staff/Child Ratio								(-)	(-)	-		+					+	
Caregiver Qualifications																		
Specialization		(+)			+		-				+	+						
Years of Education	(+)				-						-	(-)				+		
Previous Day Care Experience											+							
Experience in Current Center																		+

[1] Results noted were significant at $p<.05$; results significant at $.05<p<.15$ s.

free play. There also was a weak tendency for older children to attend less often to adults during adult-directed activity. Older children also tended to engage in more structured activities during the observations.

Individual coefficients for the covariables are not reported in the regression tables. Their contribution to the R^2 is indicated, however.

Group Composition Variables

REFLECTION/INNOVATION (Table 4.3). In both contexts--free play and adult-directed activity--more reflection/innovation on the part of children was associated with smaller groups, though the relationships were only marginally significant. In neither context was the amount of reflection/innovation related to staff/child ratio.

VERBAL INITIATIVE (Table 4.4). Children more often offered opinions in smaller groups, regardless of the activity context. The staff/child ratio in the classroom was not related to the amount of verbal initiative, however.

COOPERATION (Table 4.5). In both free play and adult-directed activities, more cooperation was observed in smaller classrooms. Amount of cooperation was not related to staff/child ratio.

NONINVOLVEMENT (Table 4.6). The level of child noninvolvement during free play activities was related to group size: noninvolvement tended to be more frequent in larger classrooms. In the context of adult-directed activities, child noninvolvement was not related to group size. There was, however, some hint of a relationship with ratio; more noninvolvement was observed in lower ratio classrooms.

AIMLESS WANDERING (Table 4.7). The frequency of aimless wandering was related both to group size and to staff/child ratio. Wandering children were more frequently observed in larger classrooms and in classrooms with lower staff/child ratios. This pattern held for free play and adult-directed activities, although the group size effect for wandering was not as strong during the adult-directed activities.

TASK PERSISTENCE (Table 4.8). Children remained involved in tasks longer where staff/child ratio was higher during both free play and adult-directed activities. A tendency toward longer activities in larger groups was found for adult-directed activities.

ORIENTATION TO ADULTS (Table 4.9). The frequency of children's orientation to adults during both free play and adult-directed activities was related to group size. Children in smaller classrooms were more often oriented toward adults. Ratio was not related to amount of attention to adults during these activities.

ORIENTATION TO CHILDREN (Table 4.10). No relationships were found between children's attention to other children and either group size or staff/child ratio.

ORIENTATION TO GROUPS (Table 4.11). During both free play and adult-directed activities, children spent more time interacting in groups when the total population of the classroom was large. Staff/child ratio showed no relationship to group orientation during adult-directed activities, and a paradoxical relationship, marginally significant at best, during free play: children attended to groups more often in higher ratio classrooms.

Summary of Group Size and Ratio Effects

For the measures shown in Table 4.12, group size was consistently related to child behavior even when other variables correlated with group size are included in the regression model. Group size effects were persistent for both free play activities and adult-directed activities, but were slightly stronger for free play. In general, the data suggest that the smaller group is a more engaging environment for young children, with higher levels of involvement in activities, reflection/innovation, verbal initiative, cooperation, and orientation to adults, and lower levels of wandering.

Staff/child ratio was rarely related to child behavior in either activity context. However, higher ratios were associated with less wandering and with greater task persistence during adult-directed activities. Although high ratios appear to have less pervasive effects than small groups, the observed relationships suggest a somewhat positive influence for higher ratios.

Caregiver Qualifications

REFLECTION/INNOVATION (Table 4.3). The amount of reflection/innovation in the classroom during free play activities was not related to caregivers' qualifications. In the adult-directed activities, specialized caregiver education/training was positively related to REFLECTION/INNOVATION.

VERBAL INITIATIVE (Table 4.4). There were no significant effects of caregiver qualifications on the frequency of children's verbal initiative.

COOPERATION (Table 4.5). The amount of cooperation observed during free play activities was related to

caregiver education/training in a child-related field. More cooperation was associated with higher proportion of caregivers with specialized training. Years of caregiver education showed significant negative relationship in the regressions for free play activities, but this effect is potentially an artifact since the simple correlation of education and COOPERATION was essentially zero. None of the caregiver qualifications was associated with cooperation during adult-directed activities.

NONINVOLVEMENT (Table 4.6). The level of noninvolvement in a classroom was negatively related to the caregiver's education/training in a child-related area. That is, there tended to be more activity in classrooms where more caregivers had specialized preparation. This result held for free play activities only.

AIMLESS WANDERING (Table 4.7). None of the qualifications variables was significantly related to aimless wandering during either free play or adult-directed activities.

TASK PERSISTENCE (Table 4.8). Children remained in activities longer where more staff had specialized preparation during both free play and adult-directed activities. For free play activities, groups where staff had more experience in other day care centers exhibited more task persistence. As in the case of cooperation, caregiver education was a significant negative regressor (for both free play and adult-directed activities) but was not strongly correlated with TASK PERSISTENCE; therefore this analytic result is questionable.

ORIENTATION TO ADULTS (Table 4.9). None of the measures of caregiver qualifications was associated with amount of ORIENTATION TO ADULTS, regardless of the activity context.

ORIENTATION TO CHILDREN (Table 4.10). During adult-directed activity periods, children in groups where staff had more years of education spent more time attending to other children. No other significant relationship was found between ORIENTATION TO CHILDREN and staff qualifications.

ORIENTATION TO GROUPS (Table 4.11). None of the measures of caregiver qualifications was significantly correlated with ORIENTATION TO GROUPS, in either context. However, in the regression analysis for adult-directed activities, experience in current day care center showed a significant negative association with ORIENTATION TO GROUPS. This effect may be an artifact of confoundings among the policy variables, given the absence of a significant first-order correlation.

Summary of Caregiver Qualifications Effects

None of the caregiver qualifications had powerful or pervasive effects (Table 4.12). However, the positive effects of the caregiver's preparation in a child-related field was seen relatively clearly. Classrooms with high proportions of staff having child-related preparation were marked by fewer uninvolved children, and more reflection/innovation, cooperation, and task persistence. Classes with highly educated caregivers also were marked by relatively high frequencies of reflection/innovation on the part of children, but also by low frequencies of cooperation and less task persistence (although these last results are possible regression artifacts).

The two experience variables showed few significant relationships to child behavior. A positive relationship between previous day care experience and TASK PERSISTENCE in free play suggested a positive influence of experience.

Experience in current center showed only one questionable association with a dependent measure, orientation towards groups during adult-directed activities.

Fall/Spring Comparisons

Associations between group size and child behavior were almost invariably consistent in direction in fall and spring; however, effects tended to be stronger in the spring and more pervasive in the sense of more often obtaining in both adult-directed and free play activities. The effects for ratio in the fall were consistent in meaning with spring effects; that is, higher ratios were associated with positive child behaviors; however, there was little overlap between the sets of dependent measures to which ratio was related in fall and spring.

The associations of years of education and specialized education/training with child behavior were also generally consistent between fall and spring. Years of education had its strongest effects on the free play behavior in the fall and on adult-directed behavior in the spring, however. The experience variables had scattered effects at both timepoints, involving different variable sets, but there were no contradictions in effect.

Determinants of Rare but Important Events

Some of the CFI codes that occurred infrequently (e.g., only a few times per thousand frames of observation) might be viewed as having unusual psychological importance or as being unusually revealing regarding the behavioral climate of a day care center. Relevant codes, termed "critical incidents," are listed in Table 4.13, along with their frequencies of occurrence in fall 1976 and spring 1977. Because of their rarity and because a code that was

Table 4.13

FREQUENCIES OF CRITICAL INCIDENTS CODES
AS PERCENTAGE OF ALL CODES

	Fall 1976	Spring 1977
Offers sympathy	0.1	0.0
Shares, helps	0.6	0.6
Receives praise	0.4	0.3
Asks for comfort	0.1	0.0
Receives comfort	0.3	0.3
Crying	0.2	0.2
Avoids, withdraws	0.1	0.1
Isolates self	0.7	0.1
Hostile exchange	0.1	0.1
Intrudes hostilely	0.2	0.1
Receives hostile intrusion	0.1	0.1
Receives rejection	0.1	0.1
Refuses to comply	0.3	0.2
Hostilely asserts rights	0.1	0.1
Temper tantrum	0.0	0.1
Receives threats	0.4	0.3
Receives physical punishment	0.0	0.0
Experiences accident	0.1	0.0

recorded once tended to recur over several frames, these events exhibited skewed distributions across classrooms, with many classes showing no occurrences of a given behavior, and other classes showing small flurries of critical events (e.g., a brief hostile exchange between children, followed by a few minutes of crying).

Ordinary regression embodies distributional assumptions that are violated by rare events of this kind. However, logit analysis, an alternative form of regression, is designed to handle such events. In essence, logit analysis estimates the odds of a rare event occurring at all in a given classroom, characterized by a given configuration of policy variables. (In contrast, ordinary regression as it has been used elsewhere predicts the frequency of a given event as a function of policy variables.)

A series of logit analyses was conducted, using as dependent variables the eighteen rare codes listed in Table 4.13 and using as independent variables the following: staff/child ratio, group size, staff education and staff experience, and two covariables--child age and staff age. Analyses were conducted separately for fall and spring, and for the Altanta Public Schools and each of the 49-center sites. Thus, for each pairing of an independent variable with a dependent variable (108 such pairs in all) there were eight separate opportunities for a positive or negative relationship to appear (four sets of centers at two different time points).

Needless to say, the pattern of outcomes is exceedingly complex if examined in detail. Relatively few relationships achieve conventional levels of statistical significance taken in isolation. However, the primary concern was not with relationships occurring in a particular place at a particular time but with broader relationships

that were fundamentally invariant across places and times. To identify such relationships, the following (admittedly somewhat arbitrary) criteria for declaring the existence of "consistent" effects were adopted.

(1) The signs of coefficients were consistently positive (or negative) in all, or in all but one, of the possible cases; and

(2) Either the inconsistent coefficient was not significant at the .05 level, or at least one of the consistent coefficients was significant at the .05 level.

Table 4.14 summarizes the results of applying these criteria to the array of data generated by the multiple logit analyses. The table is in the form of a matrix of dependent variables (rare codes) crossed by independent variables (policy variables and covariables). Wherever a "+" sign appears in a cell at the intersection of a particular dependent or independent variable, it indicates that a consistent positive association was found, by the definition above. A "-" sign, analogously, indicates a consistent negative relationship. An asterisk in a cell indicates that at least one coefficient was significant at the .05 level. (For technical reasons, logit analyses were not possible in all eight cases for every variable. Numbers listed in the right-hand column of Table 4.14 indicate the number of analyses on which each consistency judgment is based.)

Though this method of assessing consistency is approximate at best, the results are revealing. Large groups are associated with indices of conflict (hostile exchange, intrudes hostilely, receives physical punishment, receives threats, receives hostile intrusion, and of withdrawal (attends self). In only one case (receives praise)

Table 4.14

RELATIONSHIPS BETWEEN POLICY VARIABLES AND CRITICAL INCIDENTS[a]

	Child Age	S/C Ratio	Group Size	Staff Educ.	Staff Exper	Staff Age	N[b]
--Offers sympathy							8
--Shares, helps							0
--Receives praise	+	+	+	+	-	+	3
--Asks for comfort	-						8
--Receives comfort		+		-			4
--Crying				-			8
--Avoids, withdraws		+*		+			7
--Attends self			+				7
--Hostile exchange			+*				8
--Intrudes hostilely	+		+	-			6
--Receives hostile intrusion			+	-*			8
--Receives rejection	+*			-			8
--Refuses to comply		+		+			6
--Hostilely asserts rights							7
--Temper tantrum							7
--Receives threats		+	+				4
--Receives physical punishment			+*				6
--Accident							8

[a]Cell entries--"+" of "-" signs--indicate directions of consistent relationships.

[b]Numbers listed in right-hand columns indicate the number of analyses on which each consistency judgment is based.

*Indicates significance at the .05 level in at least one case.

are large groups associated with a critical event that would generally be regarded as positive. High staff/child ratios are associated with two categories of experience that might be regarded as beneficial to children (receives comfort, receives praise), but also with other categories that might be seen as negative (receives threats, avoids, withdraws and refuses to comply). High levels of staff education are associated with low likelihood of conflict and rejection and high likelihood of praise, but also high likelihood of avoidance/withdrawal and refusal to comply. Once again, group size is associated with a pattern of outcomes that, in our view, is more consistently desirable than the patterns associated with any other policy variables. In contrast, high staff/child ratios seem to be associated with a general intensification of emotional relationships; that is, with relatively extreme expressions of both warmth and anger. The highly educated caregiver appears to have a distinctive style, marked by avoidance of conflict. Unfortunately, because the critical incident analysis was pursued independently of other portions of the NDCS, no attempt was made to separate effects of education from those of specialization in a child-related field.

Child Behavior in Structured Situations

It has been mentioned that some behaviors of psychological interest occur infrequently in natural settings because of a simple lack of opportunity for children to act in ways that meet the definitions of relevant observation codes. Historically this has been one major reason why so much developmental research takes place in contrived laboratory settings. The legitimate intent of this kind of research has been to achieve maximum control over relevant variables--standardization of situations to which all subjects are exposed, and exclusion of extraneous influences of various kinds. To achieve such control, ecological validity has often been sacrificed.

The NDCS in general pursued a different strategy, attempting to maximize ecological validity at the risk of introducing many variables and great complexity into our analyses. However, in both fall 1976 and spring 1977 same-sex, same-age pairs of children were placed in two contrived situations, intended to present clear opportunities for certain types of behavior that were relatively rare in natural settings and that--if influenced by the policy variables--would represent important domains of effects. The situations provide the opportunity, but not necessity, for voluntary cooperation and sharing, and for creative and cooperative use of materials. The two structured situations were arranged as follows.

- In the limited resources situation, the children were given a Play-Doh game with one Play-Doh mold but an abundant quantity of Play-Doh. The crux of the situation was that only one child could use the mold.

- In the abundant resources situation, the children were given a Fisher-Price Play Family Village and associated materials. This toy permits independent play, cooperative play, and mutual fantasy play.

In both cases, behavior was recorded using the standard CFI. The structured situations achieved their goal of altering the frequencies of certain important forms of behavior (see Table 4.12). For example, frequencies of open-ended, cooperative play, innovative use of materials and reflective behavior all increased dramatically. However, regression analyses of selected CFI codes against six policy variables, plus age and SES covariables, revealed only scattered and, in our eyes, uninterpretable effects for the policy variables, and rather consistent and strong effects for age and SES. Older children generally engaged in much more active interchange than did younger children,

and, interestingly, low-SES children engaged in less discussion but more innovative, contemplative and problem-solving behavior than high-SES children.

For NDCS purposes the important conclusion to be drawn from this set of results is that effects of the policy variables are very much tied to the classroom situation. In more-or-less standardized situations CFI measures tend to reflect powerful and enduring influences of general developmental status and family background. When used in natural group settings it captures group dynamics that are subject to influence by certain regulatable center characteristics.

CHAPTER FIVE: THE CHILD: DEVELOPMENTAL TESTS*

In addition to the behavioral observations discussed in previous chapters, the NDCS explored a variety of standardized tests and rating systems in an attempt to measure the effects of the policy variables on children's cognitive and socioemotional development. Efforts were made to find valid, reliable, practical measures of a wide range of traits and skills that have received attention in the literature of developmental psychology--not only intellectual and linguistic skills, but also interpersonal skills and dispositions (such as dependency, aggression and self-control) and aspects of cognitive style (such as reflectivity, curiosity and task persistence). However, early results indicated that, except for a few rather traditional measures of school-related knowledge and cognitive skills, available measures were not satisfactory on psychometric grounds, at least when administered under NDCS field conditions. At the same time, the study was shifting its focus away from socioemotional traits toward the day-to-day dynamics of children's groups, discussed in Chapter Four. The study's explorations of various trait measures were chronicled in the testing contractor's report at the end of Phase I[5] and in the NDCS Second Annual Report.[6]

Three tests--the Preschool Inventory (PSI), the Peabody Picture Vocabulary Test (PPVT) and a test of fine and gross motor skills that was developed for the study

---

*Most of the material in this chapter is based on the work of Robert Goodrich, NDCS Research Director, and Judith Singer.[1] Material relating to the Atlanta Public School Study is based on work by Nancy Goodrich.[2,3] Psychometric analyses of the test battery were performed by William Bache.[4]

by SRI--were used in Phase III, along with a set of rating scales--the Pupil Observation Checklist (POCL)--which describe the child's behavior in the test situation. However, the motor scales and POCL were dropped because of psychometric flaws and unpromising preliminary results, to be described later. The PSI and PPVT were the only indices of individual development used to any significant extent in Phase III investigations of the correlates of the policy variables.

Critics, including some consultants to the NDCS, have questioned the use of these tests on the grounds that they are culturally biased and fail to address many important developmental goals of day care, particularly those concerning social and emotional growth.[7] However, inclusion of these tests in the NDCS measurement battery can be justified. Although tests like the PSI and PPVT admittedly measure knowledge and skills that are more readily available to white, middle-class children than to poor and/or minority children--and are therefore inappropriate measures of intelligence or general cognitive skill--the tests do to some degree predict success in school. Preparing the child for school is an important function of day care in the view of both parents and providers.[8] Mastering specific skills and knowledge is only one part of school readiness, but it is an important part. Thus, tests that measure selected, school-relevant skills play a legitimate role in measuring the outcomes of day care, so long as they are not the sole or primary measures used. As stressed earlier, NDCS test results were interpreted in the context of data from natural observations; the study's conclusions rest on a broad pattern of findings, not on results from tests alone.

Dependent measures used in NDCS analyses were not raw test scores at a single time point, but measures of change from fall to spring. Careful attention was paid

to well-known technical problems that arise in measuring change, and novel approaches to dealing with these problems were developed. Fall-to-spring changes in children's performance on the PSI and PPVT proved to be responsive to variations in regulatable center characteristics, notably group size and the education or training of caregivers in fields related to young children.

Procedures and Instruments

The Phase III test battery was administered to 1383 children in October 1976 and to 1061 children in April-May 1977. Only children tested in the fall were retained in the spring sample. Tests were conducted in the 57 study centers by testers recruited on site and trained by SRI. (Details of the recruitment and training process are provided in SRI's Phase III report.[9]) Tests were administered individually, over a two-day period. On the first day, the PSI was administered, and the POCL was completed by the tester; on the second day, the PPVT and motor scales were administered, and the POCL was again completed by the tester. Descriptions of the four instruments follow; however, it should be borne in mind that the NDCS analyses focused only on the PSI and the PPVT.

The Preschool Inventory (PSI)

Developed by Bettye Caldwell for the Educational Testing Service, the PSI has demonstrated its reliability and sensitivity to center- and home-based intervention in several large-scale studies such as the Head Start Longitudinal Study, the Head Start Planned Variation Study and the National Home Start Evaluation. The PSI is an inventory of the skills and knowledge presumed to be relevant for the preschool child's future success in school. Most of the items are verbal.

Some of the areas of knowledge covered by the test include colors, shapes, sizes and spatial relationships (e.g., the child's understanding of prepositions such as "under," "over," and "in"). (For a full description, see the handbook prepared by the Educational Testing Service.)

The PSI was designed as a measure of school readiness, not as a test of general intelligence. Unlike IQ tests, scoring involves no correction for age. A child's score is simply the number of items correct* and is highly sensitive to age and to the child's family background. Thus the test makes no pretense of "culture fairness." It is frankly intended to assess the child's preparation for a school system shaped and dominated by America's majority population--the white middle class. However, available evidence suggests that the PSI predicts school success even for children who are neither white nor middle class. In the Head Start Longitudinal Study, children's PSI scores, measured at age four, were significant predictors of children's achievement on third-grade tests of math and reading, as well as on the Raven Colored Progressive Matrices, a measure of perceptual problem-solving ability. A correlation of .59 was reported for the achievement scores and of .64 for the Raven test.[12] In addition, the PSI correlates with the Stanford-Binet, itself a predictor of school success.

A 64-item version of the PSI was administered during Phase II of the NDCS. Subsequent analyses of these

---

*During Phase II the NDCS experimented with a scoring system recommended by Hertzig et al.[10] and used by SRI in the Head Start Planned Variation Study,[11] in which incorrect answers are distinguished from failures or refusals to answer. The system is designed to reduce bias due to the child's unfamiliarity or discomfort with the test situation --a state which presumably leads to nonresponse. However, because the overwhelming majority of errors were wrong answers rather than nonresponses, Hertzig-Birch scoring was dropped.

data indicated that shortening the test entailed little sacrifice of information and also would free time to add the PPVT to the test battery. The correlation between the short (32-item) and long (64-item) versions was .96; therefore the shorter version was used in Phase III. The internal consistency (alpha) of the Phase III test was .84, compared to .90 for Phase I. Fall-to-spring stability (i.e., the fall/spring correlation) was .77, compared with .87 for the longer test. Paradoxically, these results were not altogether encouraging in view of the plan to measure gains in test scores during Phase III. As pointed out by Stanley,[13] high stability can be a drawback in measuring change. However, subsequent analyses, dscribed in detail later in this chapter, showed that reliable and meaningful change scores could be constructed for the PSI.

The Peabody Picture Vocabulary Test (PPVT)

The PPVT was included in the Phase III battery to provide an explicit measure of language skills. The PPVT is a measure of receptive vocabulary; on the test the child is asked to choose which of several pictures matches a stimulus word that is read aloud. Widely used in developmental research, the PPVT has consistently shown high reliability and has correlated well with measures of scholastic achievement and ability.[14]

The version of the PPVT used in Phase III differed from the original test in two important respects. First, SRI used revised pictures, modified by the Educational Testing Service (ETS) for use in the Head Start Longitudinal Study.[15] The ETS revision was intended to reduce cultural bias in the test by increasing the number of black persons in the illustrations and by diversifying the roles they

represent. (The original PPVT contained only two black figures--a Pullman porter and an African native.) Second, the version of the test used in the NDCS contained 90 items, rather than the 150 in the original. The 150 items on the original test are arranged in ascending order of difficulty, with later items appropriate for children older than preschool age. SRI pretested the first 60 items on a preschool population similar to that of the NDCS and found both floor and ceiling effects. SRI therefore dropped items 1-10 and included items 61-100 to increase variability at both ends of the scale.

The PPVT showed excellent inter-item homogeneity and high stability over time. Inter-item consistency (alpha) was .96. The fall-to-spring test-retest correlation was .80. Subsequent investigation, described later, indicated that the test would support change score analysis. PPVT scores were highly correlated with PSI scores (r = .74 in the fall testing period). Some of this correlation was due to the fact that scores on both tests increase with age; however, even with age controlled the partial correlation between the two tests was .64.

Children's gains on the two tests from fall to spring were less highly correlated (r = .39). Results reported later suggest that the determinants of change in the PSI and PPVT are somewhat different, although changes in both measures proved sensitive to regulatable center characteristics.

SRI Fine and Gross Motor Tests

SRI created two brief tests, one of fine and one of gross motor skills, using items common to many standardized tests such as the McCarthy Scales of Children's

Abilities and the Denver Developmental Screening Test. The fine motor items required the child to:

1. copy a circle
2. copy a plus sign
3. draw a person (six body parts)
4. build a tower of eight blocks
5. build a bridge with blocks

The gross motor items required the child to:

1. balance on one foot for ten seconds
2. jump in place
3. jump over the width of a sheet of paper
4. take two hops on one foot
5. walk forward heel-to-toe four steps
6. walk backward heel-to-toe four steps
7. catch a bounced ball three times

Separate fine and gross motor scores were obtained from the two tests. Phase III psychometric data showed that the meaning of these scores was clouded by both ceiling effects and low reliability. Nevertheless, gain scores were constructed by the procedure outlined later; the psychometric properties of the gain scores were explored, and some initial effects analyses were performed. The gain scores, averaged to center level, had relatively modest reliabilities (.36 and .45 for the fine and gross motor scales, respectively), comparable to some of the observation-based measures discussed in previous chapters. While these modest reliabilites were not in themselves sufficient reason to discard the motor scales, preliminary effects analyses gave no hint of relationships between motor gains and regulatable center characteristics; therefore the analysis was not pursued further.

Pupil Observation Checklist (POCL)

The POCL consists of nine five-point scales designed to assess the following bipolar dimensions of child behavior:

1. resistive - cooperative
2. shy - sociable
3. outgoing - withdrawn
4. involved - indifferent
5. defensive - agreeable
6. active - passive
7. gives up - keeps trying
8. quiet - talkative
9. attentive - inattentive

In Phase II, items on the POCL tended to cluster into two groups. Children's ratings on items 1, 4, 5, 7 and 9 tended to vary together, suggesting an underlying dimension of task orientation. Similarly, items 2, 3, 6 and 8 varied together, suggesting an underlying dimension of sociability. This clustering, which occurred in both the fall and spring, duplicated a similar clustering in the POCL data from the National Home Start Evaluation.[16] Thus, the POCL appeared to tap two important dimensions of behavior rather consistently.

The names of the POCL items suggest traits of children. However, POCL ratings were not made by adults who knew the children well, but by SRI testers. Thus the POCL is best viewed as an indicator of the child's state during testing and not as a measure of enduring traits of sociability or task orientation. As noted by Irving Sigel (personal communication), comfort in a test situation (or, more generally, comfort with strange adults) is itself a useful trait for children about to enter the school system. Sigel's persuasive argument and the clearcut structure

exhibited by the test items led to a preliminary decision to retain the POCL in the Phase III battery.

However, in Phase III, task orientation ratings showed a pronounced ceiling effect; fully 40 percent of children received the highest possible rating in spring 1977. With so little variability in the data, potential effects of day care center characteristics on task orientation could not be detected. Sociability scores did not show such extreme ceiling effects. However, a reexamination of Phase II data showed that analyses of change in sociability would be meaningless: when different testers rated children on successive days, day-to-day rate-rerate correlations were low (r = .44, on the average). The day-to-day correlations were barely higher than rate-rerate correlations from fall to spring (r's ranged from .37 to .42 for different testing sessions). Thus an apparent change in a child's POCL score could reflect rater disagreement and general instability of behavior in the test situation, rather than genuine change. For these reasons, further analysis of POCL ratings was abandoned.

Measurement of Change

As noted in Chapter Two, the issue of change is important for measures such as the PSI and PPVT, which capture characteristics of individual children that are relatively stable over time and relatively general across situations. Unlike observation measures, these test scores cannot be construed as descriptors of classroom dynamics or atmosphere. Hence it is of little interest whether classes or centers differ in distribution of PSI or PPVT scores, or even if such differences are associated with regulatable center characteristics. Such differences or relationships might be due solely to preexisting differences in the types of children enrolled in different types of

centers, and not to effects of centers themselves. What is of interest, of course, is the effect of center characteristics on the rate of children's growth.

Measurement of change poses a host of technical problems, as pointed out by many authors.[17,18,19] Simple difference or gain scores, e.g., differences in children's scores on the PSI from fall to spring, may appear to remove the effects of entering scores, isolating that part of a child's performance that is attributable to his environment during the interval between testings. Unfortunately, this simple approach can produce misleading results.

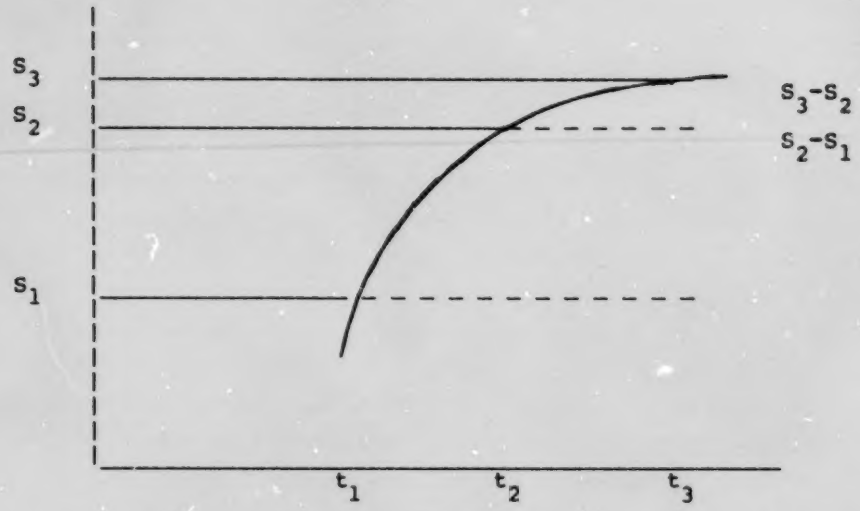
One reason for the deceptiveness of gain scores is that their reliabilities tend to be low.[20,21] What appears to be genuine change is often random measurement error, even when the test in question is relatively free of such error when applied at a single time point or two closely spaced time points (i.e., when the test is reliable in the customary sense). This problem is particularly likely to arise when scores are highly stable, that is, when persons tested at widely separated time points tend to retain their relative standings, as is true for the PSI and PPVT. Even if the underlying trait or skill being measured is perfectly stable for everyone in the tested population, scores for the same individuals tested twice will not correlate perfectly because of measurement error; reliability thus sets a ceiling on measurable stability. Measured change always incorporates this error component, as well as any real change that may occur. If the trait or skill in question is very stable (if real change in relative standing is small), stability correlations will approach the ceiling set by test reliability, and measured change will be dominated by the error component, which will be large relative to real change, though perhaps small absolutely.

But measurement of change is problematic for additional reasons that go beyond reliability limitations. Measuring change associated with a particular day care environment, when children are changing dramatically in all environments as a function of age (or, more precisely, of the maturation and experience that inevitably accompany age) is like shooting at a target moving faster than the bullet. Or, to shift the ballistic metaphor slightly, it is as if each child is on a developmental trajectory determined by powerful forces outside day care. The child's center experience causes a perturbation in the trajectory, up or down, but the perturbation may be small relative to the motion inherent in the trajectory itself. Such perturbations may be socially and psychologically significant, despite their small relative size. However, to detect them requires that the analyst have a thorough understanding of the shape of the underlying trajectory; otherwise, serious misestimates of the center effect can result. The analytic problems inherent in this sort of situation have been explored by Bryk and Weisberg.[22]

Figure 5.1 illustrates this general point with a specific analytic issue that confronted the NDCS. The typical PSI growth trajectory is curvilinear and negatively accelerated; it rises steeply at first and then flattens out. That is, young children make large gains on the test within a given time interval, while older children make smaller gains within the same interval. Similarly, children whose initial scores are relatively high tend to gain less in a given time interval than children whose initial scores are lower. In the figure, the time intervals t1-t2 and t2-t3 are equal. The child's PSI score rises rapidly (from S1 to S2 during the first interval, when the child is relatively young and begins with a relatively low initial score. In the second interval, when the child is older and begins with a higher initial score, the score rises only

Figure 5.1

TYPICAL PSI GROWTH CURVE



$t_2 - t_1 = t_3 - t_2$

from S2 to S3. (For clarity, the actual degree of curvilinearity is exaggerated in the figure.)

This pattern implies that average gain scores might vary from center to center because of differences in age composition or distributions of initial scores. If these differences were also associated with the policy variables (e.g., if centers serving younger children tend to have smaller groups or high staff/child ratios) spurious associations might be found between these policy variables and PSI gains. A traditional approach to dealing with this problem has been to use post-test scores as dependent variables in multiple regression and to use pretest scores (along with other background covariables) in the regression model, in effect removing variance attributable to these factors in order to isolate the effects of the explanatory variables of interest. However, Bryk and Weisberg, among others, have shown that this approach in many cases fails to compensate adequately for the nonindependence of entering status and subsequent gains.

Robert Goodrich, Research Director of the NDCS, conducted a thorough investigation of this issue and succeeded in devising "generalized change scores" that had the desired property of independence from entering scores and age. Goodrich approached the problem from three different angles. First, he devised adjusted gain scores specifically to meet one criterion implied above--scores whose expected covariance with age would be zero. Second, he used the more traditional method of regressing spring ("post-test") scores on fall ("pretest") scores, correcting the coefficient for measurement error (Lord-Porter Correction) and treating the residuals (deviations from the regression line) as estimates of change in children's relative standings, adjusted for entering scores. Finally, he applied modeling techniques borrowed from engineering systems theory to data

from a group of 110 children who had been tested at four time points--fall and spring of both Phase II and III. The resulting model predicts an individual's score at t+1 from his or her score at time t, adjusted by several factors that are either fixed for the population or vary randomly with a distribution whose parameters are fixed for the population. By rearranging terms within the model, Goodrich identified a particular form of adjusted change score from t to t+1 that had a constant expected value for the population and in particular was independent of a child's age and pretest score.

These three different methods produced very similar results. All three techniques yield generalized change scores of the same very simple form:

$$\text{Generalized Change Score} = S_{t+1} - KS_t$$

where: S_t = individual child's score when tested at time t

S_{t+1} = individual child's score at the next test occasion, time t+1

K = a constant less than one

Estimates of the adjustor coefficient K derived by the three different techniques are quite close to one another: .88 for the age covariance method, .86 for the traditional regression approach and .91 $\pm$ .05 for the longitudinal modeling techniques. The age covariance coefficient, .88, was used in all PSI analyses.*

---

*Detailed discussions of techniques for constructing generalized change scores appear in Volume IV of the NDCS final report[23] and in a paper presented by Robert Goodrich at the 1979 meetings of the American Educational Research Association.[24]

For the PPVT, the longitudinal method of change score calculation could not be used because the test was administered in Phase III only. However, the age covariance technique was used, yielding a generalized change score of the form $S_{t+1} - .88\ S_t$, identical to the form for the PSI (by coincidence).

Properties of Generalized Change Scores

As indicated in Chapter Two, generalized change scores had moderately high generalizabilities when averaged to the center level--the level of aggregation at which analyses were to be conducted. The center-level generalizability of PSI gains was .63 and for the PPVT was .58.*

In addition, generalized change scores proved to have two other properties that were important for their analysis and interpretation. They were essentially unaffected by race, socioeconomic status and previous day care experience of children, as well as other background characteristics that might have been confounding factors in investigations of the effects of the policy variables. However, they were strongly associated with specific patterns of family behavior, indicating their sensitivity to the climate of adult-child interaction.

Effects of General Background Characteristics

Background characteristics--genetic endowment, family influences, previous day care experience and a host of other factors--presumably affect the absolute level of a child's performance on tests such as the PSI and PPVT. But do these factors affect generalized change scores? The

*As shown in the next section, two child-level covariables were associated with PPVT gains. Adjustments were therefore made to the gain scores, and after adjustment, the generalizability of the gain scores was .53.

answer is not obvious. On one hand, generalized change scores were constructed so as to be independent of the child's starting point or pretest score. If pretest scores fully summarize the past effects of background factors and predict their future effects, generalized change scores should be unrelated to background factors. On the other hand, background factors might show "emergent" effects during the pretest/post-test interval (i.e., effects independent of those contained in the pretest score).

To address this issue, generalized PSI change scores were regressed against a set of ten background variables, including the child's race, age, sex, the total amount of time the child had been in the center as of January 1, 1977, and six family descriptor variables--family income, mother's education, number of people in the home, number of adults in the home, number of siblings and number of children under age 12 in the home. Data were drawn from 687 children--all of the children tested in both fall and spring of Phase III for whom all necessary background data were available. The ten background variables together accounted for only six-tenths of one percent of the variation in PSI gains. This finding had an important analytic consequence: it implied that investigations of the effects of the policy variables would not need to make use of any of these general background characteristics as child-level covariables. In other words, children's fall-to-spring gains would not require further adjustment, beyond the correction for the pretest score shown in the above equation, to compensate for confounding effects of income, education, race, previous day care experience and so forth.

Findings for the PPVT were similar but not identical. A parallel analysis using the same ten regressor variables and the same 687 children showed that two variables

--race and number of adults in the home--explained about 2.2 percent of the variance in PPVT gains. (The contribution of other variables was negligible.) Although the effects of background variables were minor, PPVT gains were adjusted to the account of their contribution. Analyses of the effects of center characteristics were performed using both the adjusted and unadjusted generalized PPVT change scores. Virtually no difference in policy conclusions resulted from the adjustment, but effects were generally weaker for the adjusted scores, as discussed in a later section.

Effects of Family Process Variables

A subsidiary investigation was conducted for the longitudinal sample of 110 children who had been tested at four time points. For these children, additional background data, supplementing the information discussed in the previous section, were available. Derived from interviews with parents in Phase II,[25] the data covered a variety of parental childrearing practices and attitudes. Several of the interview questions had previously been shown by Virginia Shipman and her colleagues[26] to relate to children's test performance.*

Regression analysis at the child level showed that four "family process" variables drawn from Shipman's questions were strongly associated with generalized PSI change scores.

---

*Abt Associates is indebted to Virginia Shipman of the Educational Testing Service for permission to use several of the questions devised for the ETS-Head Start Longitudinal Study and for her help in selecting the questions to be used.

Fully 30 percent of the variance in gains was attributable to these four: (1) family takes newspaper; (2) child has specific favorite story; (3) child spends time with father--all positively related to gains--and (4) number of adults other than mother who watch television with child--negatively related to gains. With family income controlled, "child has favorite story" became nonsignificant, but the remaining three variables continued to account for 27 percent of gain score variance, independent of income. When five of the background variables discussed in the previous section were controlled, "family takes newspaper" still accounted uniquely for 11 percent of gain score variation. The latter analysis represented overcontrol; it yielded an extremely conservative estimate of the proportion of variance attributable to family process variables, independent of status indicators such as income, education or race. The true proportion lay somewhere betwen 11 percent and the uncontrolled value of 30 percent. And of course, this was the proportion explained by proxy variables such as "family takes newspaper," or "child spends time with father," which obviously represent complicated patterns of parent-child interaction, rather than explaining children's cognitive gains in themselves. Presumably, more extensive and refined measurement of interactions could be expected to boost the amount of variance explained.

This subsidiary analysis was of interest primarily because it implied that generalized gain scores were highly sensitive to variations in adult-child interaction. Taken together with results reported in the previous section, the findings suggested that relevant patterns of interaction vary within racial and socioeconomic groups far more than they vary between groups. However, the analysis was not sufficiently refined to specify the most effective forms of interaction. Moreover, because relevant data were available for such a small sample (only two children per center on

average), the family process measures could not be used as covariables in estimating the effects of center characteristics.

Center-to-Center Differences

Given that PSI gains, at least, are sensitive to environmental influences in the home, the question arises whether the day care center also has an important effect. Do gains on the PSI, the PPVT, or both, vary systematically from center to center, or are they essentially random across centers, dependent wholly on the powerful effects of the home environment? How large are differences from center to center, and how significant in a practical sense? How reliably are centers characterized by high or low gains? (These questions of course apply to NDCS behavioral data as well as test scores, but only the test scores allowed comparison of child-to-child differences with differences produced by the centers.)

These are important questions for policy because current regulations are usually enforced at the center level. Centers rather than particular classrooms or caregivers are declared eligible or ineligible to serve federally funded children. In effect this enforcement policy assumes that staff/child ratio, group size, staff qualifications and so forth are center characteristics, varying more from center to center than from classroom to classroom within centers (an assumption shown in the generalizability analyses in Chapter Two to be largely but not entirely correct). The policy also implicitly assumes that quality varies more across centers than within centers. The correctness of this assumption depends on the answers to the questions posed in the preceding paragraph. If center-to-center differences in particular measures of quality (e.g., gain scores) are substantial and reliable, the assumption is correct, at least for these measures. It then becomes reasonable to

dissect these differences further, asking what portion is due to center-to-center variation in staff/child ratio, to group size, and so forth. On the other hand, if the differences are minor or unreliable, the assumption is incorrect and further center-level analysis is pointless, although comparisons at other levels (e.g., the classroom) might succeed.

To determine the magnitude of center-to-center differences in gain scores, the total child-to-child variation was partitioned into a portion attributable to centers and a portion attributable to differences among children within centers and to measurement error. The partitioning was accomplished by a series of one-way analyses of variance, each using one of the gain scores as a dependent variable, and using the 57 centers as "levels" of a single, independent classificatory variable. (A random effects analytic model, discussed by Graybill[27] as "Model V," was used. This analysis treats accidental center-to-center differences, such as would arise if children were assigned randomly to otherwise identical centers, as error variance and not as part of the systematic variation between centers--as would occur in a fixed-effects analysis of variance or a regression using centers as a set of dummy variables.) The results of this analysis are summarized in Table 5.1 and presented in more detail by Goodrich and Singer[28]. As shown in the first row of the table, about 9 percent of total child-to-child variation in PSI gains and 8 percent of variation in PPVT gains is attributable to the center that the child attends. These center effects are highly significant in the statistical sense, that is, extremely unlikely to be due to chance. Thus there <u>are</u> systematic, measurable differences in gains from center to center.

Table 5.1

CENTER CONTRIBUTION TO VARIANCE
IN GENERALIZED CHANGE SCORES

	Generalized Change Score		
	PSI	PPVT (unadjusted)	PPVT (adjusted)
Percent of Variance Due to Center	9.3%	8.2%	7.5%
Significance of Center Effect	<.001	<.001	<.001
Estimated Standard Deviation of (True) Center Mean	1.14	2.30	2.18

Are the center-to-center differences large enough to be important in any practical sense? Answering this question is partly a matter of statistics and partly a matter of judgment. Given that the proportion of variance in test score gains attributable to centers is less than ten percent, many laymen and some researchers might be tempted to conclude that the center effect is minor. However, the practical meaning of "explained variance" is not intuitively obvious. If some dependent measure varies enough, or is important enough, accounting for even a tiny fraction of its variance may be a major practical achievement.

The third row of Table 5.1 represents a step toward translating the variance figure into more intuitive terms. The row exhibits a set of center-to-center standard deviations, which may be taken as estimates of expected or typical differences between random pairs of centers. (Any

particular pair, of course, could show larger or smaller differences.) The estimates reflect "true" center impact, free of measurement error. (Measurement error increases variability of center means, so that the standard deviation of measured means exceeds the standard deviation of true means.)

For the PSI any two centers typically differ by a little more than a point in true gains over the six-month period from fall to spring of Phase III. The average fall-to-spring generalized PSI change score was 6.3 points, or 1.05 points per month. Thus the typical center difference of 1.14 points represents about 1.1 months difference in growth over a six-month period, or a difference in growth rate of about 18 percent. For the PPVT, the typical difference between centers is somewhat over two points for both the adjusted and unadjusted measures. The average adjusted PPVT gain was 7.8 points, or 1.3 points per month. Thus the typical center-to-center difference of 2.18 points represents a difference of 1.7 months growth over a six-month period, or a 28 percent difference in growth rate. In the judgment of the study's staff, these center-to-center differences are developmentally significant, especially when viewed in the context of the observational data which tend to vary in parallel with gain scores.

Center-Level Results: The 57-Center Pooled Sample

Presentation of test results in this chapter is organized differently from the preceding chapters on observations of caregivers and children. Instead of treating the 49-Center and Atlanta Public school (APS) samples separately, this chapter first discusses center-level findings based on all 57 Phase III centers as a group, and then breaks out the APS sample for investigation at the classroom level.

The principal reason for poolng all 57 centers was to increase statistical power. Pooling was not necessary for class-level analyses of observation data, since the number of classes was relatively large; however it was helpful in center-level analyses, for which degrees of freedom were fewer. (As indicated in Chapter Two, class-level analyses of gain scores were not possible in the 49-center study because enrollments in many classes shifted significantly from fall to spring; the relative stability of the APS classes, however, allowed class-level analyses to be conducted for this study. Child-level analysis was ruled out on mathematical grounds.) Pooling of center-level data from both studies was justified because the experimental treatments had almost no effects on gain scores. In any case, results for the 49-center sample proved to be essentially similar to those for the pooled sample, as shown later.

The 57-center analysis is based on a total of 896 children for the PSI and 845 for the PPVT. Because of missing data, the numbers are smaller than the group of 1061 previously mentioned as being tested at both time points: not all children tested at both time points were administered both tests on both occasions.*

Three sets of independent variables were used in the center-level analysis: classroom composition variables (staff/child ratio, group size and number of staff)**; caregiver qualifications variables (years of education,

---

*There is no evidence that the children included in the analysis differed on important background variables from these children without complete test data.

**On the basis of preliminary results, logged values of the composition variables were used in most analyses including all of those reported below.

highest degree achieved, presence or absence of education or training in a child-related field, previous day care experience and experience in current center); and a set of covariables (center averages for mother's education, family income, number of adults in the home, fraction of children in the center who were white, a poverty index describing the neighborhood surrounding the center, and the time intervals between administrations of the PSI and PPVT).

Measures of classroom composition and staff qualifications were discussed in Chapter One. In the gain score analyses classroom composition measures were based on observations averaged over the year.* Averaged observations describe the child's environment during the entire interval between tests, and thus it seemed appropriate to examine them in relationship to gain scores, which presumably reflect gradual changes in relatively long-lasting characteristics of the child. In this regard, gain scores are in marked contrast to observed behavior. Behavioral observations were used to describe the group dynamics of the classroom at a point in time and, as shown in previous chapters, were responsive to more proximate measures of the policy variables.

The covariables listed above require some explanation. Earlier it was shown that covariables at the child level (e.g., background variables such as previous day care experience, race and family income) have little or no effect on gain scores. However, as several methodologists, notably

*Only observations for the morning hours (9:00-12:00) were included in these yearly averages, because classrooms were most stable in this period and because, in most centers, educational activities were concentrated in these hours.

Cronbach,[29] have pointed out, such variables have different meanings at individual and aggregate levels, and the two kinds of effects must be considered separately. For example, the effect on a child of his or her own family's income must be distinguished from the effect on a child of the average family income level of all children in the center the child attends. When averaged to the center level, income becomes a kind of "contextual" variable. The income level of the center may well have an effect on a child's gains, even when his own family income does not, or vice versa. Hence it was necessary to explore the effects of several contextual variables, namely center averages of mother's education, family income and number of adults in the home, as well as the racial composition (measured by the fraction of white children) of the center.

Three additional center-level covariables were explored. One, a poverty index, was the fraction of families in the census tract surrounding the center with incomes below the poverty line. The poverty index, like the ecological variables constructed by averaging scores of individual children, was a measure of the socioeconomic climate of the center. The other two covariables were simply measures of the interval between administrations of the PSI and PPVT. These intervals varied from center to center, with differences ranging up to a full month. Because gain scores are directly dependent on the interval between tests, it was necessary to determine whether center-to-center variations in the intertest interval were distorting the pattern of center-mean gains.

Results for the covariables can be summarized briefly: they had no important effects themselves, and their inclusion in regression models had little or no effect on regression coefficients or *t*-statistics obtained for the policy variables. To simplify the findings presented below,

covariables will generally be ignored, and models investigating only various combinations of policy variables will be discussed in detail.

Also, virtually no interaction effects attributable to combinations of policy variables were detected. Therefore the discussion concentrates entirely on main effects.

PSI Regression Results: Overall

A preliminary regression including all policy variables and covariables suggested that four of these regressors were related to PSI gains. In order of the strengths of their relationsips to gain scores, these were group size, proportion of caregivers with child-related education/training, caregiver experience in current center, and previous day care experience. All other variables were nonsignificant. Inspection of scatterplots and correlations reinforced the impression that group size, child-related education/training and previous experience were important, but the picture for the other significant regressor--tenure in current center--was less clear. Accordingly, a series of investigations was conducted to verify and clarify the relationships between PSI gains and the four most promising policy variables. In addition, despite the fact that preliminary analysis gave no sign that staff/child ratio or years of education were related to PSI gains, these variables were also investigated further because of their potential policy importance.

In one analysis, results of which are shown in Table 5.2, the most powerful of the classroom composition variables--group size--was included in a regression model along with the three qualifications variables that had initially appeared to be significant. Results showed that

Table 5.2

RESULTS OF ORDINARY LEAST SQUARES AND BIWEIGHTED REGRESSIONS OF PSI GAINS[a] ON SELECTED POLICY VARIABLES

Center-Level: n=57

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Biweighted Least Squares Coefficient	Simple Correlation	Total R^2
Observed Group Size	-3.74	-2.66	.01	-3.67	-.33	.11
Observed Group Size	-3.82	-2.82	.008	-3.58	-.33	
Previous Day Care Experience	.16	2.30	.03	.15	+.30	.19
Observed Group Size	-3.89	-2.95	.006	-3.03	-.33	
Previous Day Care Experience	.12	1.74	.09	.12	+.30	
Child-Related Education/Training	1.22	2.08	.05	1.28	+.26	.25
Observed Group Size	-4.16	-3.06	.005	-2.44	-.33	
Previous Day Care Experience	.18	2.47	.02	.18	+.30	
Child-Related Education/Training	1.96	3.17	.003	2.11	+.26	
Experience in Current Center	- .17	-1.33	.19	- .23	-.09	.31

[a]PSI Gains are generalized change scores averaged to center level.

the effects of group size were significant and stable regardless of which qualifications variables were entered. (The negative coefficient indicates that higher gains were found in smaller groups.) Previous day care experience and child-related education/training showed fairly consistent positive relations to PSI gains, but the relative strengths of these relationships varied somewhat, depending on which other qualifications variables were entered. Tenure in current center was not significant when other variables were entered, suggesting that its emergence in the preliminary regression may have been artifactual.

Biweighting, which corrects for outlier effects, did not alter this picture. However, the biweighting process singled out three outlier centers, which were deleted from a subsequent set of analyses. In addition, regressions run with centers weighted according to the number of children tested in each suggested that small centers, where only a few children were tested, had exerted a disproportionate and somewhat distorting influence on the unweighted results. Accordingly, weighting was used in these further analyses. Principal results of regresssions, based on the reduced sample of 54 centers, weighted by number of children tested, appear in Table 5.3.

Results shown in the table reinforce the conclusions already suggested: centers that maintain small groups have higher mean gains on the PSI than centers that maintain larger groups. Centers where a high proportion of staff have child-related education/training or large amounts of previous experience in day care, also show higher gains than other centers. When parallel regressions were run with staff/child ratio in place of group size, not only did ratio show no relationship to gain scores, but the relationships shown by the qualifications variables weakened to the point

Table 5.3

RESULTS OF WEIGHTED AND WEIGHTED-BIWEIGHTED REGRESSIONS OF PSI GAINS[a] SELECTED VARIABLES

(Center-Level; n=54)

Policy Variables	Weighted Regression Coefficient	t	Significance of t	Biweighted-Weighted Regression Coefficient	Simple Correlation	R^2 (for weighted regression)
Group Size	-3.79	-2.74	.009	-3.40	-.33	.13
Group Size	-3.81	-2.84	.008	-3.38	-.33	
Previous Day Care Experience	.16	2.02	.05	.15	+.30	.19
Group Size	-4.31	-3.24	.002	-3.13	-.33	
Child-Related Education/Training	1.35	2.55	.02	1.57	+.26	.23

[a]PSI Gains are generalized change scores averaged to center level.

of nonsignificance. When group size and ratio were both included in models, alone or in conjunction with qualifications variables, group size was consistently linked to PSI gains while staff/ child ratio was not. Exploration of models including years of education revealed that this variable was related to PSI gains only occasionally and only when child-related education/ training (with which education is moderately correlated) was omitted. Thus formal education per se, independent of child-related content, seemed to make no contribution to children's gains on the PSI.

The stability of these results was examined in several ways. First, biweighting was used to compensate for distortions due to outliers. As Table 5.3 shows, biweighted coefficients for group size were fairly close to the least squares coefficients and remained quite stable as other variables were introduced, implying that outlier effects (after removing three centers) were minor and, again, that group size effects were robust. Second, the covariables were reintroduced into the regresions shown in Table 5.3. Not only were the covariables themselves nonsignificant, but they exerted little or no influence on the coefficients for group size, specialization and previous experience. Third, to guard against the possibility that center-mean gain scores might be unduly influenced by extreme individual scores within a center, all regressions were re-run using median rather than mean center-level change scores as dependent variables. The results were weaker than those shown in Table 5.3 but followed the same pattern.

PSI Regression Results: Subsamples

With tests as with observation data, subsample analyses were designed to serve as a type of cross-validation of the main findings and to indicate whether the effects of

the policy variables differ across sites, center types and populations served. Replication of the main results in most or all subsamples would rule out any possibility that the results were due to a few extreme centers or to confounding of regulated center characteristics with geographic region, center auspices and fundings, or socioeconomic characteristics of children.

In one set of subsample analyses, centers were divided according to their auspices (public versus private) their primary funding source (federal versus nonfederal) race of children served (predominantly black versus predominantly white) and income level of families served (above versus below the sample median of $6,000 in 1976). A simple summary regression of PSI gains against group size, previous day care experience of caregivers, and proportion of staff with specialized child-care education was run for each of these subsamples. Resulting coefficients and significance levels appear in Table 5.4.

Effects of group size and child-related education/training are stronger and more significant in public centers than in private centers and in centers serving mostly black children than in centers serving mostly white children. Effects of group size are also stronger and more significant in centers serving children from low-income families than in centers serving middle-income groups and in federally funded centers than in non-federally funded centers. The effects of previous day care experience are uniformly nonsignificant when the sample is partitioned, suggesting that this particular effect may lack robustness. (This issue is discussed further below.)

Results shown in the table are potentially important for federal policy. On the whole, relationships of regulatable center characteristics to test scores appear to

Table 5.4

REGRESSION COEFFICIENTS FOR PSI GAINS[a] AGAINST THREE POLICY VARIABLES, BY AUSPICES, FUNDING SOURCE, RACE AND INCOME

(Unweighted Center-Level Regressions; n=57)

	Group Size	Experience	Child-Related Education/Training
All	-4.29**	.20	1.29*
Auspices			
Public	-4.96**	-.10	1.70*
Private	-3.16	.40	1.50
Funding			
Federal	-5.49**	1.11	.41
Nonfederal	-3.26	-.12	1.61
Race			
Black	-6.47**	.22	1.81*
White	- .64	.27	.86
Income			
Above Md	-3.93*	-.22	1.97
Below Md	-5.22**	.96	.44

*p<.05
**p<.01

[a]PSI Gains are generalized change scores averaged to center level.

be strongest for centers serving the low-income, publicly subsidized children at whom policy is particularly directed. This finding may indicate that experiences in day care affect the test performance of those children more than that of middle class children, white children and children in parent-fee centers--a more advantaged group whose home environment may offset center effects. Or, the finding may merely indicate greater variability and/or different patterns of correlation among characteristics of centers serving the poor, compared to centers serving other populations. In any case, the finding suggests that group size and specialization are especially powerful regulatory levers for the federal policymaker who is concerned primarily with Title XX care. Carrying this line of interpretation still further, the results might be used as justification for federal regulations per se, which are intended in part to provide federally supported children with developmental benefits beyond the minimum guaranted for all children by state licensing requirements.

A second set of subsample analyses focused on site and regional differences and similarities. The sample was partitioned into four sections: Atlanta Public School centers, Atlanta centers other than those operated by the public schools, Detroit centers and Seattle centers. None of these subsamples included enough centers to support separate statistical analyses. Consequently, analyses were carried out by deleting one subsample at a time and re-runing the final set of regressions discussed earlier within the reduced sample. Following this step, subsamples were deleted two at a time, leaving pairs as reduced sample. Not all possible pairs were examined; rather an attempt was made to select pairs most likely to produce results discrepant from those of the 57-center analysis, in order to subject the 57-center results to the most severe test possible and to highlight differences that might exist between subsamples. Outcomes of this analysis appear in Table 5.5.

Table 5.5

RESULTS OF REGRESSIONS OF PSI GAINS[a] ON SELECTED POLICY VARIABLES

(Regressions Weighted by Number of Children in Center with Valid Gain Scores; n=54)

Sites:	Group Size	Previous Day Care Experience	Group Size	Special-ization	Staff/ Child Ratio	Previous Day Care Experience	Staff/ Child Ratio	Special-ization
All (OLS Coefficient)	-3.81	.155	-4.31	1.35	1.53	.135	1.98	1.04
(t-statistic)	(-2.84)	(2.02)	(-3.24)	(2.55)	(.890)	(1.60)	(1.19)	(1.84)
(Biweighted Coefficient)	-3.38	.153	-3.13	1.57	-5.17	.158	.335	1.87
APS, Atlanta-NonAPS, Detroit	-4.02	.140	-4.84	1.48	1.34	.136	1.85	1.13
	(-2.44)	(1.50)	(-3.03)	(2.46)	(0.62)	(1.31)	(0.900)	(1.72)
	-3.65	1.37	-3.26	1.73	-0.22	1.55	0.12	2.07
APS, Detroit, Seattle	-4.59	.155	-5.46	0.98	6.45	.030	6.09	.301
	(-3.10)	(0.98)	(-3.13)	(1.41)	(2.00)	(.171)	(1.78)	(.359)
	-5.38	.157	-5.29	1.01	6.02	.046	4.03	.939
APS, Atlanta-NonAPS, Seattle	-3.74	.161	-4.68	1.62	1.60	.147	2.5	1.24
	(-2.29)	(2.02)	(-2.83)	(2.28)	(.904)	(1.71)	(1.44)	(1.67)
	-2.41	.151	-1.72	2.17	-.851	.152	.066	.582
Atlanta-NonAPS, Detroit, Seattle	-2.15	.154	-2.60	1.34	-.395	.148	.508	1.11
(49-Center Study)	(-1.75)	(2.36)	(-2.08)	(2.49)	(-2.09)	(2.15)	(.369)	(2.00)
	-1.58	1.56	-.153	1.86	-1.60	.181	.091	1.78
APS, Atlanta-NonAPS	-3.99	.142	-5.73	1.93	1.27	.147	2.56	1.40
	(-1.81)	(1.38)	(-2.62)	(2.15)	(.539)	(1.33)	(1.09)	(1.46)
	-1.35	.130	-2.34	2.29	-1.31	1.50	.435	2.22
Detroit, Seattle	-3.08	.115	-2.81	0.38	.537	.025	.533	.101
	(-1.79)	(0.75)	(-1.68)	(0.45)	(.184)	(.157)	(.188)	(.110)
	-3.02	.116	-2.75	0.42	.609	.035	.497	2.22
Atlanta-NonAPS, Seattle	-1.22	.162	-1.78	1.46	-.150	.158	1.35	1.67
	(-0.89)	(2.76)	(-1.24)	(2.35)	(-.117)	(2.48)	(1.03)	(2.48)
	-0.52	.164	-0.96	2.17	-3.34	.181	.687	2.37
APS, Detroit	-6.49	.034	-6.62	1.12	11.86	-.150	11.06	-0.84
	(-2.54)	(.119)	(-2.72)	(1.26)	(2.01)	(-.467)	(1.67)	(-.070)
	-6.76	.016	-6.78	1.15	12.03	-.130	10.50	.174

[a]PSI gains are generalized change scores averaged to center level.

On the whole the subsamples behave roughly like the 57-center sample: In the majority of reduced samples group size, child-related education/training and previous day care experience are associated with PSI gains, but staff/child ratio is not. None of the major effects is reversed in any subsample, though the effects become marginally significant, or even clearly nonsignificant, as statistical power is lost.

While it is encouraging to find no blatant contradictions of the overall results in reduced sample, the generalization that there is agreement between the parts and the whole must be qualified: (1) the group size effect is strongest in the reduced sample consisting of APS and Detroit centers, while effects of caregiver qualifications are weak. In the complementary sample consisting of Seattle and Atlanta non-APS centers, qualifications effects are strong, while group size effects are weak. (2) Positive effects for staff/child ratio are found in some reduced samples, particuarly when the Atlanta non-APS and/or Seattle centers are removed. A negative ratio effect is found in one case. (3) Effects of both specialization and previous day care experience are much diminished whenever the Atlanta non-APS center are deleted. With all Atlanta centers removed, the effect disappears altogether.

Findings (1) and (2) suggest that reduced samples consisting of Detroit and APS centers on the one hand, and Seattle and Atlanta non-APS centers on the other, are different, with effects of classroom composition (both group size and ratio) predominant in the former, and staff qualficiations predominant in the latter. Subsequent analysis has shown that these two samples do not differ in variabilities of group size or qualifications (a difference that could have produced the observed results). There are however, subtle differences in patterns of correlation among the policy

variables, and there are also differences in racial and socioeconomic characteristics that may contribute to the results. Neither of these possible explanations has been pursued far enough to be put forward with any confidence. At present, all that can be said with certainty is that it is possible, with effort, to put together reduced samples of centers within which either group size or staff qualifications have no effects. This fact does not undermine the broader conclusion that both kinds of policy variables do have effects in most samples. Moreover, samples in which the effects disappear do not correspond to sites or regions; thus there is no direct basis in findings (1) and (2) for an argument in favor of regulation at the state or regional, rather than federal, level.

Finding (3) is easier to explain. Moreover, its explanation points to an important qualification of the study's conclusions regarding previous day care experience and sheds further light on its conclusions regarding child-related education/training. Finding (3)--that experience and specialization have measurable effects only when one or both sets of Atlanta centers appear in the subsample--appears to be due in part to the fact that the centers with highest mean levels of experience and highest proportions of caregivers with child-related specialization are found in Atlanta. Deletion of the Atlanta centers diminishes the variability of experience and specialization, hence weakening their effects.

Pursuing this observation further, it was discovered that the apparent effects of experience could be traced entirely to four centers which had high PSI gains and staff with extremely high levels of previous experience. Deletion of these centers from the full sample left a nonsignificant experience effect among the remaining 53 centers. This state of affairs poses a dilemma. The four centers are not

outliers in the usual sense. They lie near (in fact, are responsible for) the regression line that best describes the relationship between PSI gains and staff experience in the 57-center sample. They show normal effects for other policy variables. To delete these centers (7% of the sample) may be tantamount to throwing away valuable information, namely that only very large amounts of experience have measurable effects on PSI gains. On the other hand, to draw such a conclusion on the basis of information from four centers is risky. In the absence of strong supporting evidence from the observation data, the effects of previous day care experience cannot be regarded as definitively established. (This issue is discussed further in conjunction with the classroom-level APS results, below.)

The high proportion of Atlanta caregivers with child-related education/training may be traced to two sources: (1) the large number of "group leaders" (lead teachers) who held Associate's degrees in child care from Atlanta Area Technical School (AAT), and (2) the state requirement that all caregivers, including aides, complete two 30-hour courses in child care ("Basics I and II"), offered by AAT, within at least three years of center employment. It is significant that PSI gains were so closely linked to child-related education/training in Atlanta, where one institution was responsible for virtually all such education and where many caregivers who had such preparation otherwise lacked formal education beyond high school (i.e., those who had taken only Basics I and II). Unfortunately, it proved impossible to separate the effects of the Basics courses from those of the more extensive two-year course at AAT. Nevertheless, the strong

relationship between PSI gains and child-related education/ training found in Atlanta may suggest that practical courses in child care, even when taken by persons with little formal education, can make caregivers more effective teachers.

PPVT Regression Results

Analyses of the PPVT paralleled those of the PSI but were less extensive. Both adjusted and unadjusted generalized gain scores were used as dependent variables; however, most attention was focused on the former.

Scatterplots and first-order correlations suggested that gains on the PPVT, like gains on the PSI, were associated with group size. PPVT gains also seemed to be associated (negatively) with number of staff, which is highly correlated with group size (but which showed a closer relationship to PPVT gains than it had to PSI gains). The most striking difference between the two tests to emerge in the first-order correlation matrix was their different patterns of association with qualifications variables. Whereas PSI gains were moderately correlated with specialization and previous day care experience, PPVT gains were correlated with years of education.

Exploratory ordinary least squares and biweighted regressions were run using as regressors (a) either group size or number of staff, and (b) either years of education or highest degree achieved. Both group size and number of staff were consistently significant or near-significant in these regressions, with number of staff a slightly more powerful predictor. When the caregiver education variables were entered, biweighted coefficients became extremely unstable, indicating distortion due to outliers. Further inspection of the data revealed that three centers were consistently atypical (i.e., had large residuals and received low weights in the biweighting process). These were the

same three centers that had been deleted from PSI analyses because of atypicality. When they were deleted from the PPVT analyses, the effects of education--which had appeared in the correlation table but proved unstable in regressions--disappeared almost entirely.

Following these exploratory analyses, new sets of least squares and biweighted regressions were estimated. Center-mean values of dependent and independent variables were weighted by the number of children tested in each center. Independent variables included group size and number of staff, each taken separately and in conjunction with each of the qualifications variables. Results appear in Table 5.6.

The most obvious feature of the table is the small proportion of variance in PPVT gains that is explained by the policy variables, in comparison to their effects on the PSI (compare Tables 5.5 and 5.2). Generally, however, effects are in the same direction, suggesting that PPVT results are best viewed as confirming stronger findings based on the PSI. Smaller groups and fewer staff are both associated with higher PPVT gains, though the association approaches significance only when number of staff is used as a regressor, accompanied by previous experience. Previous experience has the highest t-statistic of any of the qualifications variables, again due to the same four centers that produced a significant effect of experience on PSI gains. Staff/child ratio, specialization, years of education and highest degree achieved all show no hint of significant effects on PPVT gains.

Although the impact of policy variables on PPVT gains is weak overall, it is substantially stronger in the subset of centers most relevant for policy--federally funded centers serving low-income children, many of them black.

Table 5.6

RESULTS OF WEIGHTED AND WEIGHTED-BIWEIGHTED REGRESSIONS OF PPVT GAINS SELECTED POLICY VARIABLES

(Center-Level; n=54)

Dependent Variable	Policy Variables	Weighted Regression Coefficient	t	Significance of t	Biweighted-Weighted Regression Coefficient	R^2 (for Weighted Regression
PPVT Gains (Unadjusted)	Observed Group Size	-5.20	-1.86	.07	-8.59	.06
	Number of Staff	-4.84	-1.94	.06	-5.72	.07
	Number of Staff	-5.54	-2.21	.04	-6.53	
	Previous Day Care Experience	.24	1.52	.15	.29	.11
	Number of Staff	-4.83	-1.90	.07	-6.30	
	Child-Related Education/ Training	- .03	- .03	-	.92	.07
	Number of Staff	-5.24	-2.10	.05	-6.56	
	Years of Education	.45	1.28	-	.50	.10
	Number of Staff	-5.13	-2.08	.05	-6.36	
	Highest Degree Achieved	1.33	1.54	.14	1.40	.11
PPVT Gains (Adjusted)	Group Size	-4.09	-1.55	.14	-7.43	.04
	Number of Staff	-3.36	-1.42	.17	-3.38	.04
	Number of Staff	-4.06	-1.72	.10	-4.27	
	Previous Day Care Experience	.25	1.63	.11	.26	.08
	Number of Staff	-3.65	-1.52	.14	-4.08	
	Child-Related Education/ Training	.22	.66	-	.28	.05
	Number of Staff	-3.56	-1.49	.15	-3.77	
	Years of Education	.22	.66	-	.28	.05
	Number of Staff	-3.51	-1.48	.16	-3.71	
	Highest Degree Acheived	.68	.82	-	.75	.05

[a]The adjusted gain was calculated as:

$$GPPVT_{adjusted} = GPPVT_{unadjusted} - 1.85(\text{FRACTION WHITE}) - 1.05(\text{ADULTS IN HOME})$$

where: GPPVT = Generalized PPVT Change Score, averaged to center level.
FRACTION WHITE = Proportion of children in a center who were white.
ADULTS IN HOME = Center average of number of adults living with each child.

Centers were partitioned as described earlier by auspices, funding source, predominant race of children served and income level. Regressions of (adjusted) PPVT gains on group size, experience and specialization, parallel to those previously used for the PSI, were estimated for each subsample. The results are shown in Table 5.7. With the unaccountable exception of the public/private auspices distinction, group size has significant effects on the more policy-relevant side of each dichotomy. Experience and specialization also approach or achieve significance in several of the more policy-relevant subgroups. Again, PPVT results are loosely congruent with findings based on the PSI but must be seen as supportive rather than decisive in themselves.

Class-Level Analyses: The Atlanta Public Schools Study

In the Atlanta Public Schools (APS) study, unlike the 49-center quasi-experiment, class-level analyses of gain scores were both feasible and conceptually appropriate. The APS study was designed around class-level manipulations of caregiver education and staff/child ratio. APS classes were fairly stable throughout Phase III; few children transferred in or out. Consequently, meaningful class-level scores could be computed by averaging gain scores across all children within each class who were tested in both fall and spring. Similarly, group size, staff/child ratio and staff qualifications were stable for all APS classes, except those in one center which frequently merged classes into one large group. Thus, in the APS sample as in the 57-center sample, the policy variables were measured at class level with a reasonable degree of reliability. In all analyses reported here, classes were weighted according to the number of children tested. Analyses are based on thirty classes, the 29 included in the design shown in Chapter One, plus an

Table 5.7

REGRESSION COEFFICIENTS FOR PPVT GAINS[a] AGAINST THREE POLICY VARIABLES, BY AUSPICES, FUNDING SOURCE, RACE AND INCOME

(Unweighted Center-Level Regressions; n=57)

	Group Size	Experience	Child-Related Education/Training
All	-5.25	.65	.61
Auspices			
Public	-1.62	1.33	-.97
Private	-13.96*	-.74	6.20
Funding			
Federal	-7.35**	-.39	1.30*
Nonfederal	-4.47	-.32	2.79
Race			
Black	-8.85**	1.29*	1.16
White	- .76	-.08	-.47
Income			
Middle	-3.12	-.15	1.91
Low	-8.60*	.15	1.24

*p<.05
**p<.01

[a]PPVT Gains are generalized change scores averaged to center level.

additional class that was underenrolled in the early fall (and thus excluded from the randomized design) but filled shortly thereafter.

Dependent variables were generalized PSI and PPVT gain scores, averaged to the class level. PPVT scores were unadjusted, since race--the principal adjustor variable--was the same for all children in the APS study. Independent policy variables included staff/child ratio, group size, number of caregivers, years of education, level of education of lead teacher (referring to the three levels defined in the APS experimental design*), child-related education/training, experience in current center and, previous day care experience. In addition, the following background covariables (again, averaged to class level) were explored: mother's education, child's sex (represented as fraction of the class who were girls), family income, number of adults in the home, number of siblings, number of children under age 12 in the home, and age of next youngest sibling.

Findings from these class-level investigations confirmed results of center-level analyses in important respects, but also presented some puzzles and contradictions that have not been resolved fully. Regression results for the PSI showed a very strong relationship between group size and gain scores, one that remained strong regardless of which other policy variables are included in the model (Table 5.8). However, in contrast to the 49-center results, staff/child ratio was also related to PSI gains--alone and in conjunction with group size and child-related education/training. Child-related education/training itself was

*Level of education was a three-valued variable; however, ratio, the other experimentally manipulated variable, was treated as a continuous variable rather than categorized into two treatment levels.

Table 5.8

ATLANTA PUBLIC SCHOOL STUDY:
RESULTS OF REGRESSIONS OF PSI GAINS ON SELECTED POLICY VARIABLES

(Weighted, Class-Level; n=30)

Policy Variables	Ordinary Least Squares Coefficient	t	Significance of t	Simple Correlation	R^2
Group Size	-.31	-4.19	.001	-.62	.38
Staff/Child Ratio	28.76	2.07	.05	.36	.13
Group Size	-.29	3.93	.001	-.62	.45
Staff/Child Ratio	20.98	1.83	.08	.36	
Group Size	-.29	-4.11	.001	-.62	.51
Staff/Child Ratio	25.54	2.26	.04	.36	
Child-Related Education/ Training	3.57	1.78	.09	.15	
Group Size	-.40	-5.00	.001	-.62	.50
Experience in Current Center	.50	1.86	.08	.63	
Previous Day Care Experience	-.84	-1.89	.08	-.01	

positively related to PSI gains, although the relationship fell short of significance and was not as strong as it had been in center-level results (possibly because most APS caregivers had specialized training, restricting the variation of this independent measure).

PPVT regression results differ somewhat from those of the overall study (Table 5.9). In the APS study, previous day care experience had a strong positive relationship to PPVT gains. Group size and specialization showed relationships in the expected directions, but these did not achieve significance in the regression model. Tenure in current center also showed no relationship to PPVT gains.

The above findings were subjected to several validity checks. First, given the relatively small sample of classes (30), effects of atypical, "outlier" centers could easily distort results significantly. To test for such effects, biweighted regressions were run, resulting in no substantial change in outcomes. Second, class-level covariables were introduced into regressions along with policy variables. Age of closest sibling was found to be a significant predictor of PSI gain, and mother's education was a significant predictor of PPVT gain. The significance of these covariables was probably due to the fact that they were highly correlated with the policy variables included in the regression models for predicting gain scores (group size/age of closest sibling =-.48; staff/child ratio/age of closest sibling = .32; previous day care experience/ mother's education = .36). However, when the covariables were entered into regressions with the policy variables, the overall results did not change. Thus, the major results do not appear to be threatened by covariable effects.

In sum, the APS results confirm the conclusion of the center-level study that small groups are associated with

Table 5.9

RESULTS OF THREE REGRESSIONS OF PPVT GAINS ON SELECTED POLICY VARIABLES--APS

(Weighted, Class-Level; n=30)

Policy Variables	Ordinary Least Squares Coefficient	Standard Error of Coefficient	t	Significance of t	Simple Correlation	R^2
Previous Day Care Experience	2.31	.74	3.12	.007	.51	.26
Previous Day Care Experience	2.34	.74	3.15	.006	.51	.28
Experience in Current Center	-.40	.44	-.89	-	-.12	
Previous Day Care Experience	2.20	.79	2.77	.01	.51	.26
Group Size	-.04	.13	.26	-	-.23	
Previous Day Care Experience	2.18	.75	2.92	.009	.51	.29
Child-Related Specialization	3.87	3.54	1.09	-	.26	

high gains on the PSI. The simple correlation between group size and PPVT gains was similar in magnitude and direction to that obtained in the center-level study, but in multiple regression the effects of group size were dominated by those of previous experience, acting in conjunction with mother's education. Child-related education/training, which had shown a significant relationship to gains on the PSI in the center-level study, shows no such relationship in the APS study. However, because there was almost no variation among APS caregivers on the specialization dimension, this finding should not be seen as a failure to replicate results of the larger study. (Almost all APS caregivers had taken Basics I and II or received degrees from AAT). APS results hint that staff/child ratio may be related to gains on the PSI--a finding borne out by the results of the APS experiment, summarized in Chapter One, but not by the center-level study. The APS study also showed that previous day care experience with staff was positively related to gains on the PPVT.

Conclusions

NDCS findings on links between the impact of regulated center characteristics and children's gains on the PSI and PPVT lend themselves to a deceptively easy summary: Several of the policy variables seem to influence cognitive gains. This result is strongest for group size. Small groups are associated with more rapid gains on both tests in the 57-center study and on PSI in the class-level APS study. The magnitude of the effect is large in many cases, and it withstands virtually all tests of its validity. Child-related specialization also appears to influence cognitive gains. Its effects are confined to the PSI and are neither as large nor as pervasive as those of group size; however, they are evident in both class and center-level analyses. For previous experience, NDCS results are

less definitive: The variable shows some positive effects on both tests. However, these effects are confined to a few centers in the 57-center analysis and are confounded with an "ecological" (class-level) family background effect (of mother's education) in the APS study. Other policy variables do not appear to have consistent, important effects.

Though many qualifications and caveats could be appended to the foregoing summary, on the whole it represents a fair statement to the policymaker. It is, however, excessively mechanical. It conveys an impression that group size, for example, is a knob that can be twisted to push gain scores up or down. It ignores the processes of human interaction that link gross features of the classroom, such as group size, to a child's cognitive growth. This important connection is completed in the next chapter.

CHAPTER SIX: LINKS BETWEEN CLASSROOM PROCESS AND CHILD TEST SCORES*

This chapter explores relationships between classroom process in NDCS day care centers (the observed behavior of caregivers and children) and children's gains on standardized tests of school readiness. These exploratory analyses were intended to discover whether and how classroom process mediates the relationship between the policy variables and child outcomes--that is, the degree to which the effects of the policy variables can be traced through classroom process to children's performance. It was previously shown (in Chapters Three and Four) that both caregiver behavior and child behavior are linked to some of the policy variables. In Chapter Five, links between policy variables and children's gains on the standardized tests were reported. The remaining connection to be established is that between classroom process and test score gains.

Few educators or day care providers would argue that limiting group size or hiring caregivers with specialization in child development would automatically ensure greater cognitive gains for children in day care. Rather, it is likely that the caregiver's behavior and the response of the children in her class form essential links between the policy variables and child test scores. Caregivers who have specialized in child development behave differently in the classroom from those who have not; for example, they interact more with children in a variety of ways, and these behavioral differences are likely to contribute to increased cognitive gains. Similarly, children in smaller groups behave differently from children in larger groups; for example,

---

*The material in this chapter is based on work by Judith Singer; reported in detail in Volume IV-C of the NDCS Final Report.[1] Ms. Singer is the principal author of this chapter.

they show more creative, verbal, and intellectual activity--and their behavior is likely to influence their test scores. However, the particular behaviors most closely linked to cognitive gains, and the role played by these linkages in mediating the effects of the policy variables, remain to be determined.

All of the analyses reported in this chapter are exploratory in the sense that they were not guided by a strong theory about the specific connections of classroom process and child gains on cognitive tests. However, common-sense ideas about teaching and learning provided some hypotheses about which caregiver and child behaviors could be associated with gain scores. For example, the AFI code INSTRUCTS and the CFI code REFLECTION/INNOVATION were expected to relate to greater gains, since it seemed plausible that caregivers who spend more time in direct teaching should have children who learn more, and classrooms where more children engage in thoughtful, creative activities should show higher average gains.

The exploratory analyses were also guided by earlier findings on the relationships between behavior and the policy variables. In many cases there were significant relationships between a policy variable and a caregiver or child behavior and between the same policy variable and cognitive gains. In such cases, either the behavior or the policy variable or both might be associated with higher gains. For example, group size was a strong predictor of COOPERATES on the CFI and was also related to cognitive gains. These relationships may indicate that there exists a causal chain linking group size to cooperation to cognitive gains: children cooperate with adults more in smaller classes and children who cooperate more achieve higher cognitive gains.

If cooperation in fact wholly mediates this effect of group size in the manner indicated, it should be related to cognitive gains even when it occurs in large groups (though it occurs less frequently in such groups). Group size would not show a relationship to cognitive gains that was independent of the level of cooperation. Alternatively, the relationship between group size and cognitive gains may be mediated wholly by some other variable, such as REFLECTION/INNOVATION, or possibly by behavioral variables not measured at all. In such cases, cooperation would not show a relationship to cognitive gains independent of group size, but group size, would show a relationship independent of cooperation. Finally, cooperation might be one of several variables mediating the effects of group size, in which case both cooperation and group size would be independently associated with cognitive gains. To disentagle such rival hypotheses, a series of regression analyses were carried out, using as regressors, different combinations of policy variables and behavioral variables known to relate to the policy variables, and using test score gains as dependent variables. These analyses were undertaken with the hope of clarifying the relative roles of the policy variables and classroom processes in influencing children's cognitive gains.

Methods and Analytic Issues

Data Sources

Analyses of linkages between classroom processes and child test scores were based on data from a number of sources, all of which have been described in detail in previous chapters of this volume, and will simply be summarized here. The dependent variables in these analyses were generalized gain scores constructed for the PSI and PPVT. The classroom process data were obtained with the

two observation instruments--the Adult Focus Instrument (AFI) and the Child-Focus Instrument (CFI). Most of the independent behavioral variables used in the linkage analysis were identical to those described earlier. However, some additional variables were also constructed, primarily to strengthen the analysis statistically by capitalizing on correlations among previously discussed measures. Table 6.1 lists the AFI and CFI codes used in the analyses of linkages.

On the AFI, the variables included the major WHAT and TO WHOM codes and three macro-codes or summary variables. The MANAGEMENT macro-code is identical to that discussed in Chapter Three while the other two macro-codes differ somewhat from earlier variables. SOCIAL ACTIVITY is calculated as the difference between the previously defined macro-code, SOCIAL INTERACTION, and the individual code OBSERVES. The statistical jusification for this combination is a negative correlation between the two variables, suggesting that their combination would be a stronger variable than each code separately. Concomitant with the empirical advantage of the new macrocode is its substantive interpretation. This new SOCIAL ACTIVITY code represents the balance struck by a particular caregiver between interaction with children and passive observation of their activities.

The third AFI macro-code, GROUP SCALE, can also be rationalized statistically from the negative correlation of TO LARGE GROUP and TO MEDIUM GROUP and substantively from the notion of balance in the direction of attention of a caregiver. To what degree does she attend to large groups (frequently the whole class) as opposed to somewhat smaller groups? In some sense, GROUP SCALE can stand as a representative for the preservation of the class as a unit as opposed to its division into groups.

Table 6.1

CFI AND AFI VARIABLES USED IN THE LINKAGE ANALYSIS

CFI VARIABLES
- Individual Codes
  - VERBAL INITIATIVE
  - CONSIDERS
  - ADDS PROPS
  - WANDERS
  - RECEIVES INPUT FROM ADULT
  - TASK PERSISTENCE
  - NON-INVOLVEMENT
  - MOVES WITH PURPOSE
  - MONITORS ENVIRONMENT
  - COOPERATES
  - ATTENTION TO ADULT
  - ATTENTION TO CHILD
  - ATTENTION TO GROUP
  - ATTENTION TO ENVIRONMENT
  - OPEN ACTIVITY
  - CLOSED ACTIVITY
- Macro-Codes
  - REFLECTION/INNOVATION (Considers + Adds Prop)
  - INDIFFERENCE (Wanders - Reflection/Innovation)
  - CLASS STRUCTURE (Open Activity - Closed Activity)

AFI VARIABLES
- To Whom
  - TO STAFF
  - TO CHILD
  - TO SMALL GROUP
  - TO MEDIUM GROUP
  - TO LARGE GROUP
- What
  - COMMANDS
  - CORRECTS
  - DIRECT QUESTIONS
  - RESPONDS
  - COMFORTS
  - PRAISES
  - OBSERVES
  - INSTRUCTS
  - ADULT ACTIVITY
- Macro-Codes
  - GROUP SCALE (To Large Group - To Medium Group)
  - MANAGEMENT (Commands + Corrects)
  - SOCIAL ACTIVITY (Direct Questions + Responds + Comforts + Praises - Observes)

Analyses of the CFI utilized most of the individual micro-codes discussed in Chapter Four, together with three macro-codes. One macro-code, here termed CLASS STRUCTURE, is identical to the CLASSROOM ACTIVITY BALANCE discussed in Chapter Four; this variable represents the relative amount of children's participation in instructional vs. structured activity.) A second, REFLECTION/INNOVATION, was already discussed extensively in Chapter Four. A third, INDIFFERENCE, was constructed by subtracting the frequency of REFLECTION/INNOVATION from the frequency of AIMLESS WANDERING. Construction of this variable was justified primarily by the negative correlation between its components. For purposes of the linkage analysis--in contrast to previously reported results on the CFI alone--frequencies of codes and constructs were summed across teacher-directed and free-play activity periods in order to reduce the number of variables examined to a relatively compact set.

Unit of Analysis

Analyses linking the observation and test data were done at center level rather than class level, consistent with the other analyses of cognitive gains but different from analyses of the AFI and CFI alone. This choice was necessary because, as discussed earlier, measures of change could not be constructed at the class level without sacrificing large amounts of data and introducing various sampling biases. So many children moved from one classroom to another between the fall and spring testing that very small numbers of children would have constituted each "class," for purposes of calculating changes scores, and many children with complete data could not be assigned to a particular class. In addition, it is likely that attrition from each class would be selective in unknown ways, further undercutting the usefulness of the sample. Finally, classes are frequently organized by age of child,

so that older children are promoted to older age groups as the year goes on and younger children are admitted. Thus, the children in the NDCS sample who stayed in one class would have been younger or more immature than was true of the center as a whole, and their test scores alone would not have fairly represented the classroom or the center.

However, aggregating classroom process measures to the center level posed some problems. The AFI in particular appeared to be indicative of classroom patterns (which here are synonymous with lead-teacher patterns) as opposed to center-level patterns. By aggregating across classes within a given center, a substantial amount of the generalizability of the AFI measures was sacrified; generalizabilities fell from roughly .7 - .9 at class level to .2 at center level. (Center-level generalizability for CFI variables were approximately the same as class-level generalizabilities, so that this problem did not apply to the CFI measures.)

The choice, then, was to conduct linkage analyses at the class level, which would require the omission of test data on many children, or conduct the linkage analyses at the center level and lose information on the class-level variability of the observation data. The loss of information on variability seemed minor compared with that incurred if two hundred children's test scores were to be omitted from the analysis; therefore center-level analyses were used.

Sample

For the AFI, only data on lead teachers were used in the linkage analyses, since the data for aides were incomplete and those for teachers were more representative of the centers. For the tests, children included in the analysis had to have both a valid pretest and post-test for either the PSI or PPVT, as well as valid CFI data. As with

all test score analyses, however, it was not necessary to have valid test scores for both tests. Thus the sample of children for the PSI is slightly different than that for the PPVT. In addition, only children whose race was reported as white or black were examined; all children reporting race as "other" were omitted from analysis (less than 4% of all children). In this way, problems concerning children whose native language was not English were virtually eliminated.

Results of Analyses of Classroom Process and Children's Gain Scores

The first step employed in the process-outcome linkage was to examine two-way plots of PSI and PPVT generalized gain scores versus each of the process measures from the AFI and CFI. On the basis of these graphs, several centers were determined to be potential outliers. Second, weighted correlations were computed with and without the potential outliers, resulting in exclusion of these centers from further analysis. Regression models were then constructed to predict cognitive gain scores from various combinations of policy variables and process measures. The results of each of these analytic steps are presented below.

Preliminary Analyses: Graphs and Correlations

The two-way plots of the PSI and PPVT gain scores and the process measures suggested that the CFI data bore a strong relationship to PSI gain scores, while AFI data were more strongly associated with PPVT gain scores. (Recall that although scores on the PSI and the PPVT at a single time are highly correlated, the generalized gain scores of children on these tests are relatively independent. At the center level, the correlation between the cognitive gain scores used in the process-outcome analysis is 0.39. As a

result, variables that are significantly correlated with one of the measures are not necessarily correlated with the other measure.)

In addition to suggesting that the two tests might be associated with different types of behavioral data, these graphs showed that there were several centers that did not fit into the overall pattern for many of tne dependent and independent variables. Three of these centers were the same ones that had been set aside from the cognitive main effects analyses For the PSI, one additional center appeared to be rather atypical; for the PPVT there were two other centers that might be considered outliers. To ensure that future results would not be unduly influenced by these centers (four for the PSI and five for the PPVT), the next stage of analysis (correlations) was done with and without these centers to determine their effect upon results.

For each generalized gain score, weighted correlation matrices were constructed both with and without the outlier centers. As expected, these centers were unduly influencing results. For example, the correlation between PSI GAINS and COOPERATES is 0.19 if all centers are included in analysis; when the four atypical centers are omitted, the correlation jumps to 0.42. These four centers fell so far away from the general pattern that they made an effect that is actually quite dramatic appear to be just barely significant. Therefore, the outlier centers were set aside from subsequent analyses; only the results for the remaining centers will be discussed.*

The weighted correlations (excluding the outlier centers) reinforced the previously mentioned indication that

*These centers were included in several biweighted analyses and were found to receive very low weights, thus reinforcing the notion that they were distorting the overall correlational pattern.

the PSI is more highly associated with CFI than AFI data and the PPVT with AFI data slightly more than CFI. Children in centers where PSI gains are high show high frequencies of COOPERATION and REFLECTION/INNOVATION, and low frequencies of aimless wandering (reflected in the INDIFFERENCE variable). Moreover, individual children receive input from adults more often in these environments, there are more structured than open-ended activities, and caregivers attend more to medium- than large-sized groups. However, two anomalous findings also appear: a negative correlation between TASK PERSISTENCE and PSI GAIN ($r = -0.32$), and positive correlation between NONINVOLVEMENT and PSI GAIN ($r = 0.31$). (As will be seen shortly, these anomalous relationships were not confirmed in regression analyses, whereas other relationships suggested by the pattern of simple correlations were confirmed.) Caregiver behavior did not appear to bear a strong relationship to PSI gains. The only significant correlation was with GROUP SCALE, such that center-level gains were higher where caregivers focused more attention on medium-sized groups as opposed to large ones.

In the case of the PPVT, simple correlations suggested that the only CFI variables related to the measures of cognitive gain are those dealing with movement. Higher gains occur where children move with purpose, do not often wander aimlessly and involve themselves in reflective activities more often than they wander. In contrast to the PSI results, the PPVT gain scores show relationships to several AFI measures. In centers with large PPVT gains, lead teachers attend more frequently to individual children and more frequently to medium- than to large-sized groups. In addition, they engage in more MANAGEMENT and SOCIAL ACTIVITY with the children.

Regression Analyses

Multiple regression was used to model the combined associations of the CFI, AFI and cognitive gain scores. As described earlier, subsets of the CFI and AFI variables were entered into the analysis because many of these independent variables were multicollinear, and results from comprehensive analyses would not have been interpretable. In addition, there were just slightly more than fifty cases available for center-level analyses; yet there were almost forty independent variables of interest. Thus the number of degrees of freedom available was severely restricted, also rendering individual coefficients and R^2's all but meaningless. Of course, the problems imposed by multicollinearity and limited degrees of freedom were not averted merely by selecting small sets of regressors; selection itself creates problems of interpretation. The interpretability of the results depends on empirical and conceptual support from the various main effects analyses; again, the study's ability to "borrow strength" from multiple analyses was its best protection against the ambiguities of any analysis taken in isolation.

The simple correlations were used to guide construction of the regression models.* In the models, all two-way and three-way combinations of CFI and AFI variables were tested, initially excluding those variables that, on the basis of the simple correlations, were not related to gains.** Also, the major policy variables previously found to be significantly related to cognitive gains (group size,

---

* Regressions were weighted by the appropriate number of children. In addition, weighted-biweighted regressions were estimated. Centers previously determined to be outliers were not included.

**Process variables that had nonsignificant simple correlations were subsequently entered into regression models to further investigate their behavior. Without exception, these variables remained nonsignificant.

proportion of caregivers with child-related education/training--"specialization"--and mean years of caregiver experience) were included. Finally, covariables were initially used to control for possible confoundings of race of children in the center and SES characteristics of the center, although as in the other cognitive analyses, they were subsequently found to be nonsignificant.

The most informative of the regression models constructed for PSI gain scores are presented in Table 6.2. The models reported in the table all contain at least one CFI or AFI variable that had a significant simple correlation with PSI gains and a significant regression coefficient whose direction of effect was identical to that of the simple correlation (or there was a good reason for the difference). The regressions essentially confirm the correlational results: centers in which children more frequently engage in reflective behavior, cooperate with teachers and become involved in thoughtful tasks rather than wander tend to have higher gains on the PSI; in addition, children in classes that are more structured tend to have higher gains. The stability of the results for GROUP SIZE in every model indicates that the importance of this policy variable for PSI gains is partially independent of the study's measures of classroom process. The stability of the regression coefficients after biweighting further strengthens the validity of all of the significant findings. (Note, however, that this stability is due in part to the deletion of the four outlier centers.)

These models were constructed with the intention of describing as tersely as possible the type of day care center which facilitates higher PSI gains. Toward this end, certain CFI variables included in the models act as proxies for a whole host of variables not entered into the model but correlated with the regressors used. For example,

Table 6.2

RESULTS OF WEIGHTED AND WEIGHTED-BIWEIGHTED REGRESSIONS
DEPENDENT VARIABLE: PSI GAIN SCORE*
(n=53 Centers)

Source	Independent Variables	Weighted Regression Coefficient	t	Significance of t	Biweighted Weighted Regression Coefficient	Simple Correlation	R^2
	Group Size	-0.07	-2.25	.04	-0.07	-.36	
CFI	REFLECTION/ INNOVATION	21.89	3.22	.002	22.35	.43	
CFI	COOPERATES	6.58	3.11	.004	6.77	.42	.40
	Group Size	-0.09	-2.81	.008	-0.08	-.36	
CFI	REFLECTION/ INNOVATION	21.00	2.89	.007	21.41	.43	
	SPECIALIZATION	.99	1.91	.07	1.09	.25	.33
	Group Size	-0.07	-2.34	.03	-0.07	-.36	
CFI	COOPERATION	6.74	-3.03	.005	7.07	.42	
CFI	INDIFFERENCE	-9.15	-2.19	.04	-9.52	-.32	.33
	Group Size	-0.08	-2.78	.009	-0.08	-.36	
CFI	REFLECTION/ INNOVATION	23.58	3.36	.002	24.53	.43	
CFI	CLASS STRUCTURE	-3.16	-2.35	.03	-3.30	-.24	.35

*Only those AFI and CFI variables which acted as significant predictors ($p \leq .05$) appear on this table.

COOPERATES is correlated with degree to which children receive input from adults, the amount of structure in the class and also the proportion of time children spend focusing their attention towards other children. By the same principle, the variable RECEIVES INPUT (from adults) which is not included specifically as a regressor in Table 6.2, is indeed a characteristic of centers with higher PSI gains. Due to its correlation with many of the other variables, however, it was not found to be as strong a regressor as CLASS STRUCTURE or COOPERATES, for example, and as such was not explicitly entered into the regression models.

The same approach was employed to construct regression models for PPVT gains; the results of this analysis appear in Table 6.3. As the simple correlations indicated, many aspects of caregiver behavior are associated with higher generalized gains on the PPVT, but only one CFI variable, INDIFFERENCE, is associated (negatively) with PPVT gains. Centers with higher PPVT gains tend to be characterized by more one-to-one caregiver-child interaction. These caregivers spend more time in both MANAGEMENT (commanding and correcting) and SOCIAL ACTIVITY (more time interacting, less time passively observing). In centers with higher gains, teachers spend more time with medium-sized groups as opposed to larger ones. Also, children tend to be more actively involved in intellectual/creative activities instead of wandering around the class.

Table 6.3 shows that the coefficients for the AFI variables are rather stable in the face of variation in regression models used; coefficient estimates obtained in the more inclusive models are strikingly similar to those obtained in simpler models. The initial and biweighted coefficients in all models are remarkably similar, further strengthening the stability of these findings.* That is,

---

*As before, this stability is due in part to the deletion of the four outlier centers

Table 6.3

RESULTS OF WEIGHTED AND WEIGHTED-BIWEIGHTED REGRESSIONS
DEPENDENT VARIABLE: PPVT GAIN SCORES*
(n=52 Centers)

Source	Independent Variables	Weighted Regression Coefficient	t	Significance of t	Biweighted Weighted Regression Coefficient	Simple Correlation	R^2
AFI	GROUP SCALE	-4.07	-2.02	.05	-3.88	-.41	
AFI	SOCIAL ACTIVITY	8.63	2.71	.01	8.92	.46	.33
AFI	SOCIAL ACTIVITY	10.81	3.80	.001	11.20	.46	
CFI	INDIFFERENCE	-20.47	-2.70	.01	-20.44	-.34	.32
AFI	GROUP SCALE	-4.48	-2.24	.04	-4.64	-.41	
AFI	TO CHILD	7.11	2.39	.02	7.36	.33	
CFI	INDIFFERENCE	-17.37	-2.06	.05	-17.03	-.34	.32
AFI	GROUP SCALE	-6.02	-3.41	.002	-6.16	-.41	
AFI	MANAGEMENT	24.12	3.85	.001	24.49	.25	.41
AFI	GROUP SCALE	-5.37	-2.67	.01	-5.44	-.41	
AFI	MANAGEMENT	14.78	2.26	.03	14.88	.25	
AFI	SOCIAL ACTIVITY	6.77	2.14	.04	6.96	.46	.35
AFI	GROUP SCALE	-4.16	-2.22	.04	-3.98	-.41	
AFI	MANAGEMENT	20.49	3.30	.002	21.15	.25	
AFI	SOCIAL ACTIVITY	6.68	2.31	.03	7.25	.46	
CFI	INDIFFERENCE	-24.30	-3.29	.002	-25.63	-.34	.47

*Only those AFI and CFI variables which acted as significant predictors ($p \leq .05$) appear on this table.

although the predictor variables are correlated, it is possible to estimate their separate effects through a single model. (Note that it was not possible to include TO CHILD in this all-inclusive model because its effects and those of MANAGEMENT and SOCIAL ACTIVITY became severely attenuated. As before, however, it is important to keep in mind that even though TO CHILD is not explicitly included in most of these regression models, it is included via the two AFI macro-codes with which it is correlated.)

Summary and Discussion

In sum, many structural and behavior characteristics of day care centers are associated with children's gains on the PSI and PPVT. Although it is difficult to separate out the individual components, together they describe a center in which small numbers of children and adults interact to produce an integrated, cohesive unit.

The major finding discussed in earlier chapters has been that small groups are associated with better care for children. Analyses reported in this chapter not only support this finding, but also provide additional refinements to our understanding of why group size is an important dimension of quality care. As indicated by both AFI data and the analysis of the GROUP SCALE variable, the number of children present with one or more caregivers, measured by a total head count, effectively determines the size of the "subgroups" toward which lead caregivers typically direct their attention. As the number of children assigned to a classroom increases, the size of these subgroups increases, regardless of the prevailing staff/child ratio. That is, classes are rarely divided into smaller groups of roughly equal size, even when enough adults are present to permit such division. Rather, lead caregivers appear to supervise most or all of the children in the class at once, although

aides may occasionally take one or a few children aside for special activities. The size of the "effective sub-groupings" around the lead teacher is associated with a whole range of child behaviors and outcomes.

Centers in which caregivers typically interact with medium-sized groups as opposed to large ones have higher gains on both PSI and PPVT. Children in these centers also tend to be more involved in classroom activities and spend less time wandering about. When effective groupings are large, caregivers tend to stop interacting with children and begin to stand back and passively observe classroom activities. These behavior patterns of children and caregivers appear to mediate some but not all of the affect of group size on cognitive gains. Moreover, there is some difference between the behaviors that mediate gains on the PSI and those that mediate gains on the PPVT, although there is also significant overlap.

Interactiveness on the part of the caregiver is also an important correlate of test score gains. Centers in which caregivers are more interactive and orient themselves towards children tend to have higher cognitive gains, especially on the PPVT. Further, caregivers who stand back and observe children passively, instead of interacting with them, are found in centers with lower cognitive gains. Although the type of interaction may be either managerial (commanding and correcting) or social in nature, social interaction is the stronger predictor. In fact, the amount of social interaction bears the strongest relationship to a measure of cognitive gain of any variable examined.

In addition to total interaction, the amount of <u>one-to-one</u> interaction a caregiver displays is related to test score gains. Centers in which caregivers spend a large proportion of their time interacting with individual

children tend to have higher PPVT gain scores than centers in which caregivers tend to direct their attention to groups of children.

Children who are active and integrated into the classroom activities have higher cognitive gains on both instruments, while centers in which children spend a large proportion of their time wandering have lower gains on the average. There is a distinct pattern of child behavior characterized by such behaviors as considering, contemplating, tinkering, adding props or ideas to ongoing activities, and cooperating with others, which is not only associated with less time spent wandering, but also related to higher gains, especially on the PSI.

Finally, group size shows relationships to cognitive gains that are independent of the behaviors identified above. Behavioral mediators other than those measured in the NDCS apparently contribute to the powerful and pervasive effects of this structural variable on cognitive gains.

CHAPTER SEVEN: SUMMARY AND CONCLUSIONS

To summarize and draw conclusions from the results of a policy study as complex as the NDCS is a matter of judgment and art as much as science. There are no hard-and-fast rules for choosing which among many data sets to emphasize and which to treat as subsidiary, or for deciding when a clear but relatively isolated finding should be taken seriously and when such a finding should be dismissed as an anomaly. Clearly there are technical, objective considerations in making such decisions. For example, greater emphasis should be placed on findings from large subsamples than small ones, on findings replicated in several subsamples than on those confined to a single subsample, or on particularly strong and/or highly significant relationships than on weaker relationships or on those near the statistical margin. Emphasis also should be placed on findings that are theoretically reasonable, are plausible in light of a practical understanding of how day centers function and/or are supported by previous research. But in a study that is likely to have policy consequences, nontechnical considerations must also inevitably play a role, not only in formulating recommendations but also in choosing which results to stress and which to downplay. Thus this summary makes no pretense of being entirely value-free. It is firmly grounded in data, but it also reflects an attempt to strike a balance between a desire to guide the government in purchasing the best possible care for chidren and a desire to avoid imposing unnecessarily costly and/or ineffective restrictions on providers.

The major findings of the National Day Care Study are summarized in the Preface. They are restated here, amplified by significant details from the intervening chapters.

First, variations in regulatable characteristics of day care centers are associated with significant variations in the behavior of caregivers, the behavior of children and children's gains on selected developmental tests. In the one domain for which it was possible to compare center effects with effects of factors outside the centers--the domain of test scores--about 8-9 percent of the variation in gains was attributable to centers. "Better" centers in the sample had rates of gain that were roughly 20 percent higher than those in "less good" centers. Center effects were smaller than those associated with variations in the home environment, but they were statistically and substantively significant.

Second, of all the regulatable characteristics studied, group size showed the most pervasive pattern of associations with measures of behavior and test scores: small groups were better for children than large groups. When the total number of children in the classroom was small, lead teachers tended to spend time in various forms of social interaction with small clusters of children; when the total number of children was large, lead teachers tended to spend time in passive observation of the group as a whole. Children in small groups showed more creative, verbal/intellectual and cooperative behavior than their peers in larger groups. They were less likely to be non-participants in classroom activities, and they had higher gains on standardized tests from fall to spring.

Most of these relationships were consistent in direction across subsamples, though they varied in strength and significance. Perhaps most notably, they tended to be especially strong for low-income, black children in publicly subsidized centers. Although there were differences in strength across sites (to some degree paralleling the ethnic and socioeconomic differences

just mentioned), there was little evidence of major heterogeneity that might suggest that the effects of group size are site-specific. Moreover, there was no clear numerical point of demarcation between small, "good" groups and large, "bad" ones. Most of the study's centers maintained groups of three- and four-year olds that varied in size from 12 to 24; typically, desirable behaviors decreased in frequency by roughly 20 percent, and undesirable behaviors increased by 20 percent, as group size increased within this range.

Third, staff/child ratio was also related to some aspects of interaction in the classroom, but the correlates of this critical policy variable, the focus of much of the controversy surrounding day care regulations, were less widespread than those of group size. Ratio was most clearly related to caregiver behavior: lead caregivers in high-ratio classes (those with few children per adult) showed essentially the same pattern of behavior reported above for caregivers in small groups. (However, the confounding of ratio and group size for the lead caregiver sample made it unclear whether the behavior pattern should be attributed to ratio, group size or both.) In addition, lead caregivers in high-ratio classes spent less time in overt management of children than those in low-ratio classes. They also spent more time interacting with other adults and in other activities not directly involving children. Thus some of the "contact time" potentially available to children by virtue of high adult/child ratios was spent in other ways. High ratios were <u>not</u> associated with high frequencies of one-to-one interaction between adults and children; in fact, ratio showed few systematic relationships to the behavior of children at all. Nor was ratio related to children's test score gains, except in a few isolated instances.

The relatively modest and scattered effects of ratio must be interpreted in light not only of the (deliberately) restricted range of ratios in the sample but of the naturally occurring configurations of classrooms in the day care world. As indicated in Chapter Two, most centers in the NDCS sample maintained ratios between 1:5 and 1:9 for three- and four-year olds. While this range is highly relevant for policy (covering the spectrum from the FIDCR-mandated level for three-year olds to a level close to the maximum for preschoolers permitted by the licensing requirements of many states) it is relativley narrow in an absolute sense and therefore tends to restrict detection of ratio effects. Moreover, many high-ratio classes, particularly those where total class size is large, utilize a single lead teacher and one or more aides, who are generally assigned less responsibility for the care of children. Thus high ratios often imply a kind of dilution of adult responsibility, as well as requiring that the lead teacher divert some of her energies to managing other adults. If these interpretations are correct, they imply a weakening of the potential effectiveness of ratio as a regulatory tool for influencing classroom dynamics. They also imply that with proper training and a redefinition of the role of aides, ratio could become a more effective regulatory device and the general quality of care could be increased. However, given current staffing practices, NDCS findings suggest some shift of regulatory emphasis away from ratio toward group size, though both aspects of classroom composition deserve a place in regulations.

Finally, among the various aspects of caregiver qualifications, education or training in fields specifically related to young children emerged as the strongest correlate of caregiver behavior and children's test scores. Lead caregivers with specialized education or training played a more active role with children than those without such

preparation, and children under their supervision made relatively rapid gains on standardized tests. These relationships were most clearcut in Atlanta, where substantial numbers of caregivers received relevant education or training from a single institution. They were weaker (although still in a positive direction) or nonexistent in other sites and could not be tested in the Atlanta Public School study, where almost all caregivers had relevant preparation. However, despite their restriction to certain portions of the total sample, the effected child-related education/ training may have wider generality, which has been obscured by variations in the amount and content of such education and training available at different sites. The apparent positive effects of child-related education/ training may of course be due partly to self-selection by individuals who have sought such training rather than to the benefits of training itself. Nevertheless, the presence of such individuals in a day care classroom appears to affect the quality of the child's experience and its developmental consequences. Thus, though findings with respect to this variable are somewhat tentative, their potential importance for the well-being of children in day care, in the judgment of the study's staff, overrides the methodological caveats that surround them and justifies inclusion of some training provision in federal regulations.

Even more tentative are the findings on caregivers' experience prior to employment at their current centers. Previous experience showed only scattered relationships to behavior of caregivers and children. Relationships to test scores were found in only four centers in the 49-center study and were confined to the PPVT in the Atlanta Public Schools study. On balance, while there are clear hints of positive effects, previous experience does not appear to correlate consistently with indices of quality for children--perhaps because "years of experience" is a relatively gross

variable that fails to distinguish qualities of experience and that lumps caregivers who have become expert on the job with those who have "burned out." (Experience measured in terms of tenure in the caregiver's current center had no consistent positive or negative effects.) Consequently, the NDCS did not recommend inclusion of an experience requirement in federal standards regarding staff qualifications.

Findings with respect to formal education per se--that is, education without regard to child-related content--reveal no unequivocal positive effects. In general the correlates of years of education were few and scattered. Moreover, the few apparent relationships may be due to the socioeconomic status or other background characteristics of the caregiver rather than to benefits conferred by formal education itself. Thus the data provide no support for a regulatory requirement based on years of education or degrees achieved.

REFERENCES

PREFACE

1. Ruopp, R., Travers, J., Glantz, F., and Coelen, C. Children at the Center. Final Report of the National Day Care Study: Summary Findings and their Implications. Cambridge, Mass.: Abt Books, 1979.

2. Federal Register, March 19, 1980.

3. Other supporting volumes include Coelen, C., Glantz, F., and Calore, D. Day Care Centers in the U.S.: A National Profile 1976-1977. Cambridge, MA: Abt Books, 1978; and three volumes of Technical Appendices to the National Day Care Study. Cambridge, Mass.: Abt Associates Inc., 1980.

4. Ruopp, et al., op. cit., Appendix A. p. 231.

5. Assistant Secretary for Planning and Evaluation, Department of Health, Education and Welfare. The Appropriateness of the Federal Interagency Day Care Requirements: Report of Findings and Recommendations. Washington, D.C.: U.S. Government Printing Office, 1978.

6. Travers, J., and Ruopp, R. National Day Care Study Preliminary Findings and Their Implications Cambridge, Mass.: Abt Associates Inc., 1978.

7. See Ruopp, et al., op. cit., Chapter 8, 155, and Appendix A, 230-240.

8. Ruopp, et al., op. cit., Appendix A., 236.

CHAPTER ONE

1. Coelen, et al., op. cit.

2. Ruopp, et al., op. cit., Appendix B.

3. Phase I results are presented in Abt Associates Inc., National Day Care Study First Annual Report. Cambridge, Mass.: Abt Associates Inc., 1976 and Stallings, J., Wilcox, M. and Travers, J. Phase II Instruments for the National Day Care Cost-Effects Study: Instrument Selection and Field Testing. Menlo Park, Calif.: Stanford Research Institute, 1976.

4. Phase II results are presented in Travers, J., Coelen, C., and Ruopp, R. National Day Care Study Second Annual Report. Cambridge, Mass.: Abt Associates In., 1977.

5. Ruopp, et al., op. cit., Chapter Four. More detailed discussion appears in National Day Care Study Measurements and Methods. Final Report of the National Day Care Study, Volume IV-B. Cambridge, Mass.: Abt Associates Inc., 1980.

6. Ruopp, et al., op. cit., Chapter Five. More detailed discussion appears in National Day Care Study Measurements and Methods. Final Report of the National Day Care Study, Volume IV-B. Cambridge, Mass.: Abt Associates Inc., 1980.

7. Bache, W. "Comparing Alternative Measures of Classroom Composition in National Day Care Study Measurements and Methods. Final Report of the National Day Care Study, Volume IV-B. Cambridge, Mass.: Abt Associates Inc., 1980.

8. National Day Care Study First Annual Report, op. cit.; Stallings J., Wilcox, M. and Travers, J., op. cit.; and Travers, J., Coelen, C. and Ruopp, R., op. cit.

9. Goodson, B.D. "The Adult-Focus Observation Effects Analysis." National Day Care Study Effects Analyses. Final Report of the National Day Care Study, Volume IV-C. Cambridge, Mass.: Abt Associates Inc., 1980.

10. Connell, D.C. "The Child-Focus Observation Effects Analysis." National Day Care Study Effects Analyses. Final Report of the National Day Care Study, Volume IV-C. Cambridge, Mass.: Abt Associates Inc., 1980.

11. Goodrich, R.L., and Singer, J.D. "Cognitive Change in the NDCS." National Day Care Study Effects Analyses. Final Report of the National Day Care Study, Volume IV-C. Cambridge, Mass.: Abt Associates Inc., 1980.

12. Goodrich, N.N. "The Atlanta Public Schools Day Care Experiment." National Day Care Study Effects Analyses. Final Report of the National Day Care Study, Volume IV-C. Cambridge, Mass.: Abt Associates Inc., 1980.

CHAPTER TWO

1. For a detailed discussion of the site selection process, see Ruopp, R. National Day Care Study First Annual Report, Volume II. Cambridge, Mass.: Abt Associates Inc., 1976.

2. U.S. Bureau of the Census, Census of Population: 1970, Volume I, Characteristics of the Population. Part I, U.S. Summary; Section. I Washington, D.C.: U.S. Government Printing Office, 1973.

3. "Case Studies of the National Day Care Study Sites: Atlanta, Detroit and Seattle" in National Day Care Study Background Materials, Final Report of the National Day Care Study, Volume IV-A, Cambridge, Mass.: Abt Associates Inc., 1980.

4. U.S. Bureau of the Census, op. cit.

5. Coelen, et al., op. cit.

6. F. Mosteller and J.W. Tukey. Data Analysis and Regression. Reading, MA: Addison Wesley, 1978.

7. Goodrich and Singer, op. cit.; Connell, op. cit.; Goodson, op. cit.

8. Mosteller and Tukey, op. cit.

9. Bronfenbrenner, U. "Developmental Research, Public Policy and the Ecology of Childhood." Child Development. 45, 1974, 1-5.

10. Millett, R.A. and Mathis, A. "The National Day Care Study from the Perspective of Black Social Scientists: Reflections on Key Research Issues." National Day Care Study Background Materials, Final Report of The National Day Care Study, Volume lV-A. Cambridge, Mass.: Abt Associates Inc., 1980.

11. Medley, D.M., and Mitzel, H.E., "Measuring Classroom Behavior by Systematic Observation," in Handbook of Research on Teaching, N.L. Gage (Ed). Chicago, Ill.: Rand McNally, 1963.

12. Cronbach, L.J., Gleser, G.C., Nanda, H. and Rajaratnam, N. The Dependability of Behavior Measures: Theory of Generalizability for Scores and Profiles. New York, N.Y.: John Wiley and Sons, 1972.

13. Affholter, D.P. and Bache, W.L. "The Application of Generalizability Theory to Observation-Based Dependent and Independent Measures in the National Day Care Study." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

14. Robinson, W.S. "Ecological Correlations and the Behavior of Individuals." American Sociological Review, 15, June 1950, 351-356.

15. See, e.g., Cronbach, L.J. (with assistance of J.E. Deken and N. Webb). "Research on Classroom and Schools: Formulation of Questions, Design and Analysis." Occasional Paper, Stanford Evaluation Consortium, Stanford, Calif., July 1976.

16. Singer, J. and Goodrich, R. "Aggregation and the Unit of of Analysis in the National Day Care Study." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1970.

17. Ruopp, et al., op. cit.

CHAPTER THREE

1. Goodson, op. cit.

2. Goodson, B.D. "The Classroom Environment Study." National Day Care Study Measurements and Methods. Final Report of the National Day Care Study, Volume IV-B. Cambridge, Mass.: Abt Associates Inc., 1980.

3. Goodrich, N.N. "An Analysis of the CDA Checklist Data." National Day Care Study Measurements and Methods. Final Report of the National Day Care Study, Volume IV-B. Cambridge, Mass.: Abt Associates Inc., 1980.

4. Stallings, J. and D. Broussard. Final Report: National Day Care Cost-Effects Study: Spring 1977 (Phase III) Data Collection. SRI, July, 1977.

5. Stallings and Broussard, op. cit.

6. Goodrich, N.N. "The Effects of Day Care in Eight Atlanta Public Schools Day Care Centers." National Day Care Study Effects Analyses. Final Report of the National Day Care Study, Volume IV-C. Cambridge, Mass.: Abt Associates Inc., 1980.

7. Connell, op. cit.

8. Goodrich and Singer, op. cit.

CHAPTER FOUR

1. Connell, op. cit.

2. Prescott, E., Jones, E., Kritchevsky, S., Milich, C. and Haselhoef, E., Assessment of Child-Rearing Environments: An Ecological Approach. Pasadena, Calif.: Pacific Oaks, 1975.

3. Stallings, et al., op. cit.

4. SRI International, "Training Manual: Prescott-SRI Child Observation System." Menlo Park, Calif.: SRI International, 1976.

5. For example, Kohn, M.L. and Rosman, B.L. "A Social Competence Scale and Symptom Checklist for the Preschool Child: Factor Dimensions, Their Cross-Instrument Generality and Longitudinal Persistence." Developmental Psychology. 1972, 6, 430-444.

CHAPTER FIVE

1. Goodrich and Singer, op. cit.

2. Goodrich, N.N. "The Atlanta Public Schools Day Care Experiment." op. cit.

3. Goodrich, N.N. "The Effects of Day Care in Eight Atlanta Public Schools Day Care Centers." op. cit.

4. Bache, W.L. "A Psychometric Analysis of the National Day Care Study Phase III Child Test Battery." National Day Care Study Measurements and Methods. Final Report of the National Day Care Study, Volume IV-B. Cambridge, Mass.: Abt Associates Inc., 1980.

5. Stallings, et al., op. cit.

6. Travers, Coelen and Ruopp, op. cit.

7. Millett and Mathis, op. cit.

8. Layzer, J. "Interviews with Parents." National Day Care Study Effects Analyses. Final Report of the National Day Care Study, Volume IV-C. Cambridge, Mass.: Abt Associates Inc., 1980. Also, Goodson, B.D., "The Classroom Environment Study." op. cit.

9. Stallings and Broussard, op. cit.

10. Hertzig, M.E., Birch, H.G., Thomas, A. and Mendez, O.A. "Class and Ethnic Differences in the Responsiveness of Preschool Children to Cognitive Demands." Monograph of the Society for Research in Child Development, 1968, 33, No. 1.

11. Walker, D.R., Bane, M.J. and Bryk, A.J. The Quality of the Head Start Planned Variation Data. Cambridge, Mass.: Huron Institute, 1973.

12. Shipman, V.C., McKee, J.D. and Bridgeman, B. Disadvantaged Children and their First School Experiences: Stability and Change in Family Status, Situational, and Process Variables and their Relationship to Children's Cognitive Performance. Report to the Office of Child Development, DHEW (Grant No. H-8256). Princeton, N.J.: Educational Testing Service (ETS No. PR-75-28), September 1976.

13. Stanley, J. "Reliability." In R.L. Thorndike, (ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1971, (2nd ed.).

14. Meissner, J.A., Shipman, V.C. and Gilbert, L.E. "Technical Report 15: Peabody Picture Vocabulary Test," in V.C. Shipman, (ed.), Disadvantaged Children and their First School Experiences: ETA Longitudinal Study Technical Report Series. Princeton, N.J.: Educational Testing Service (ETS Report No. PR-72-27), 1972.

15. Ibid.

16. High/Scope Educational Research Foundation and Abt Associates Inc. Home Start Evaluation: Final Report Findings and Implications. Ypsilanti, Mich.: High/Scope Educational Research Foundation and Abt Associates Inc, 1976.

17. Bereiter, C. "Some Persistent Dilemmas in the Measurement of Change." In C.W. Harris, (ed.), Problems in Measuring Change, Madison, Wisc.: University of Wisconsin Press, 1963.

18. Cronbach, L.J. and Furby, L. "How We Should Measure 'Change' - or Should We?" Psychological Bulletin, 1970, 74, 68-80.

19. Linn, R.L. and Slinde, J.A. "The Determination of the Significance of Change Between Pre- and Posttesting Periods." Review of Educational Research, 1977, 47, 121-150.

20. Stanley, op. cit.

21. Bereiter, op. cit.

22. Bryk, A.J. and Weisberg, H.I. "Use of the Nonequivalent Control Group Design When Subjects are Growing." Psychological Bulletin, 1977, 84, 950-962.

23. Goodrich and Singer, op. cit.

24. Goodrich, R. "Measurement of Cognitive Change--A Dynamic Stochastic Approach." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979.

283

25. Layzer, op. cit.

26. Shipman, et al., op. cit.

27. Graybill, F.A. An Introduction to Linear Statistical Models, Volume I. New York, N.Y.: McGraw-Hill, 1961.

CHAPTER SIX

1. Singer, J. "Classroom Process-Child Outcome Analyses." National Day Care Study Effects Analyses. Final Report of the National Day Care Study, Volume IV-C. Cambridge, Mass.: Abt Associates Inc., 1980.